DOCTORAL THESIS

Disordered eating in pregnancy: the development and validation of a pregnancy-specific screening instrument

Bannatyne, Amy Jean

Award date:
2018

Disordered Eating in Pregnancy:

The Development and Validation of a Pregnancy-Specific Screening Instrument

Amy Jean Bannatyne

BSocSc (Psyc) (Hons I), MPsyc (Clin)

Submitted in total fulfillment of the requirements for the degree of

Doctor of Philosophy

School of Psychology

Faculty of Society and Design

Bond University

23rd February 2018

Declaration of Originality

This thesis is submitted to Bond University in fulfillment of the requirements for the degree of Doctor of Philosophy (PhD). This thesis represents my own original work toward this research degree and contains no material which has been previously submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made. All raw data and analyses have been retained and are available upon request. I certify that I have made and retained a copy of this document.

Amy Jean Bannatyne

23rd February 2018

Acknowledgements

There are so many special people who have contributed to this PhD over the past three years. First and foremost, I would like to express my sincerest appreciation to my supervisory team: Dr Peta Stapleton, Dr Bruce Watt, Dr Roger Hughes, and Dr Kristen MacKenzie-Shalders. Thank you for always encouraging, supporting, and inspiring me in my research. In particular, thank you to Peta for being an incredible mentor and for gently reminding me of my abilities when I needed it. I also wish to acknowledge Bond University for the many opportunities afforded to me, and the Australian Government for the financial support offered during this PhD via an Australian Postgraduate Award.

I would like to make special mention of Dr Liz Edwards who has been an invaluable source of support to me – more than she knows. Thank you also to Dr Mark Edwards who has provided continuous support in all my studies and career goals. To my wonderful colleagues and peers, Elyse, Ashley, Rachel, Laura, Scott, Katarina, James, and Trish – thank you for your ongoing words of encouragement. I am so grateful for our friendships. A big thank you to Elyse and Ashley for proof reading the final draft of this thesis.

This thesis would not be in its current form without the support of my strongest support network, my family. To my mum (Sue) and grandparents (Terry & Hilda), thank you for always believing in me, for encouraging me to set my goals high, and for the sacrifices you all made so I could study. My achievements really are a product of your endless love, support, and compassion. To my soon-to-be-husband (Jacques), thank you for your wholehearted support in everything I do. Without your unconditional love, patience, and belief in me, I would not be where I am today.

Lastly, to all the researchers, clinicians, advocates, and women who participated in this research – thank you so much for generously contributing your time and sharing your experiences/expertise. Without your involvement, this thesis would not have been possible.

Research Publications and Outputs

Journal Publications

Bannatyne, A., Hughes, R., Stapleton, P., Watt, B., & MacKenzie-Shalders, K. (in press). Consensus on the assessment of disordered eating in pregnancy: An international Delphi study. *Archives of Women's Mental Health.* doi:10.1007/s00737-017-0806-x

Bannatyne, A., Hughes, R., Stapleton, P., Watt, B., & MacKenzie-Shalders, K. (in press). Signs and symptoms of disordered eating in pregnancy: A Delphi consensus study. *BMC Pregnancy & Childbirth*.

Additional Publications Achieved During Candidature

Bannatyne, A., & Stapleton, P. (2017). Attitudes toward anorexia nervosa: Volitional stigma differences in a sample of pre-clinical medicine and psychology students. *Journal of Mental Health, 26,* 442-448. doi:10.3109/09638237.2016.11149801

Edwards, E. J., Bannatyne, A., & Stark, A. (2017). Twelve tips for teaching brief motivational interviewing to medical students. *Medical Teacher,* 1-6. doi:10.1080/0142159X.2017.1369503

Stapleton, P., Bannatyne, A., Chatwin, H., Urzi, K.C., Porter, B., & Sheldon, T. (2017). Secondary psychological outcomes in a controlled trial of emotional freedom techniques and cognitive behaviour therapy in the treatment of food cravings. *Complementary Therapies and Clinical Practice, 28,* 136-145. doi:10.1016/j.ctcp.2017.06.004

Bannatyne, A., & Stapleton, P. (2016). Eating disorder patient experiences of volitional stigma within the health care system and views on biogenetic framing: A qualitative perspective. *Australian Psychologist.* doi:10.1111/ap.12171

Stapleton, P., Bannatyne, A., Porter, B., Urzi, K.C., & Sheldon, T. (2016). Food for thought: A randomised controlled trial of emotional freedom techniques and cognitive

behavioural therapy in the treatment of food cravings. *Applied Psychology: Health and Well-Being, 8,* 232-257. doi:10.111/aphw.12070

Bannatyne, A., & Stapleton, P. (2016). Educating medical students about anorexia nervosa: A potential method for reducing the volitional stigma associated with the disorder. *Eating Disorders: The Journal of Treatment and Prevention, 23,* 115-133. doi:10.1080/10640266.2014.976102

Conference Presentations

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *Antenatal screening for disordered eating in pregnancy: Should it be implemented? Preliminary results of two Delphi consensus studies.* Paper presented at the 2016 Bond University Faculty of Health Sciences and Medicine Higher Degree Research Conference (November, Gold Coast).

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *What constitutes disordered eating in pregnancy? Preliminary results of a Delphi consensus study.* Paper presented at the 2016 Australia New Zealand Academy of Eating Disorders Conference (August, New Zealand).

Bannatyne, A., & Stapleton, P. (2016). *Patient experiences of volitional stigma within the healthcare system: A qualitative perspective from anorexia nervosa sufferers.* Paper presented at the 2016 Australia New Zealand Academy of Eating Disorders Conference (August, New Zealand).

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *Disordered eating and pregnancy: Key issues to address to improve screening and detection in antenatal care.* Paper presented at the 2016 Eating Disorders and Obesity Conference (May, Gold Coast).

Conference Presentations Accepted, but Unable to Attend

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *Disordered or [ab]normal eating? An international consensus study exploring the definition and assessment of disordered eating in pregnancy.* Paper accepted for the 2017 International Conference for Eating Disorders (June, Prague).

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *Disordered or [ab]normal eating? An international consensus study exploring the definition and assessment of disordered eating in pregnancy.* Paper accepted for the 2016 International Marcé Society Conference (September, Melbourne).

Poster Presentations

Bannatyne, A., Stapleton, P., Hughes, R., MacKenzie-Shalders, K., & Watt, B. (2016). *Disordered or [ab]normal eating in pregnancy?* Poster presented at the 2016 Bond University Learning and Teaching Week (September, Gold Coast). Awarded best poster presentation.

Ethics Declaration

All research conducted and reported in this thesis received approval by the Bond University Human Research Ethics Committee (BUHREC). Ethics application numbers were #15278 (Studies 1 and 2) and #15964 (Study 4). Approval letters for both applications can be found in Appendix A and E, respectively. All procedures performed in studies involving human participants were conducted in accordance with the ethical standards of BUHREC and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in each respective study; this included consent for publication.

Abstract

Optimising the mental health of women during the perinatal period, inclusive of pregnancy and the first year post-birth, has been identified as a global priority. While much research has focused on depression and related conditions, disordered eating is thought to affect a similar proportion of women. A growing body of research also suggests pregnancy may represent a period of vulnerability for the precipitation, re-emergence, or exacerbation of disordered eating; however, such symptoms are often undetected and undisclosed in pregnancy. Disordered eating in pregnancy has been linked to several negative consequences such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and other obstetric and postpartum difficulties. Eating disorder scholars and advocacy organisations have argued that antenatal care should include questions regarding a woman's body weight, eating practices/attitudes, and weight control behaviour/s during pregnancy; however, there is mixed guidance as to how this should occur, under what circumstances, and using which methods/instruments. Consequently, screening for disordered eating symptomatology in pregnancy is rare. As such, the overarching aim of this thesis was to improve the identification of disordered eating in pregnancy.

To address this aim and answer the four main research questions, a mixed methodological approach was employed with four sequential studies conducted. Studies 1 and 2 used the Delphi methodology to explore professional and consumer views on the symptom expression of disordered eating in pregnancy and how this is distinguished from pregnancy-appropriate symptomatology. Professional and consumer views on the assessment of disordered eating in antenatal care were also explored. Study 3 aimed to systematically identify and evaluate general measures of disordered eating, using standardised performance criteria, to determine their suitability for use in pregnancy. Study 4 aimed to develop a pregnancy-specific disordered eating screening instrument based on the findings of Studies 1

and 2, and then evaluate the psychometric properties of the instrument using a sample of pregnant women in Australia (N = 444). The pregnancy-specific instrument was also compared to two well-known eating disorder measures.

Studies 1 and 2 revealed strong consensus that (i) disordered eating in pregnancy is somewhat similar, yet also distinct, to the experience of disordered eating in a non-pregnant context; (ii) the delineation between disordered eating and pregnancy-appropriate symptomatology is difficult to quantify, but might be assisted using several qualitative and quantitative factors; and (iii) antenatal screening for disordered eating is imperative and should occur in a routine/universal manner using brief psychometric instruments validated for use in pregnancy. The systematic review conducted as Study 3 revealed little to no evidence to support the use of existing disordered eating measures in pregnancy, highlighting the need for a pregnancy-specific measure of disordered eating to be developed and research exploring the validity of existing self-report inventories in pregnancy to be conducted. Study 4 provided preliminary evidence that the Disordered Eating Attitudes in Pregnancy Scale (DEAPS) constitutes a valid and user-friendly instrument to assess and screen for disordered eating attitudes during pregnancy. The DEAPS demonstrated a high level of internal consistency, appropriate content validity, good construct validity, and very strong concurrent criterion-related validity.

Overall, this thesis revealed that disordered eating is a relatively common experience during pregnancy and that routine/universal screening for such symptoms might be warranted in antenatal care, similar to screening for antenatal depression and anxiety. This thesis has also provided preliminary evidence that the implementation of universal screening may be feasible using the DEAPS. While further research is required to confirm the psychometric properties of the DEAPS in additional samples and different settings, this thesis has highlighted a need for policy makers to consider the inclusion of disordered eating screening

in perinatal mental health guidelines and the importance of clinicians being educated and aware of such symptoms. Routine screening of disordered eating in pregnancy may facilitate early identification and management, contributing to a positive pregnancy experience and potentially mitigating associated morbidity and costs for mothers, infants, families, and societies. Ongoing research in this area is vital, particularly the development and evaluation of evidence-based interventions to support women experiencing disordered eating during pregnancy.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALSPAC	Avon Longitudinal Study of Parents and Children
APA	American Psychiatric Association
AN	Anorexia Nervosa
BED	Binge Eating Disorder
BMI	Body Mass Index
BN	Bulimia Nervosa
COPE	Centre of Perinatal Excellence
DEBS	Disordered Eating Behaviour Scale
DEAPS	Disordered Eating Attitudes in Pregnancy Scale
DSM-5	Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition
EAT	Eating Attitudes Test
ED	Eating Disorder
EDDS	Eating Disorder Diagnostic Scale
EDE	Eating Disorder Examination (interview)
EDE-PV	Eating Disorder Examination – Pregnancy Version
EDE-Q	Eating Disorder Examination Questionnaire
EDI	Eating Disorder Inventory
EDNOS	Eating Disorder Not Otherwise Specified
EDNOS-P	Eating Disorder Not Otherwise Specified – Purging Disorder
EPDS	Edinburgh Postnatal Depression Scale
GP	General Practitioner
GWG	Gestational Weight Gain
HDR	Higher Degree Research
IOM	Institute of Medicine

KMO	Kaiser-Meyer-Olkin Index
KR-20	Kuder Richardson-20 Index
MoBa	Norwegian Mother Baby Cohort
NEDC	National Eating Disorders Collaboration
NICE	National Institute for Health and Care Excellence
NPDI	National Perinatal Depression Initiative
OECD	Organisation for Economic Cooperation and Development
OSFED	Other Specified Feeding or Eating Disorder
PCA	Principal Components Analysis
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta Analyses
PwC	PricewaterhouseCoopers
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
ROC	Receiver Operating Curves
RQ	Research Question
SCOFF	Sick Control One Fat Food Questionnaire
SIGN	Scottish Intercollegiate Guidelines Network
STARD	Standards for Reporting Diagnostic Accuracy
UK	United Kingdom
UN	United Nations
US	United States
USFED	Unspecified Feeding or Eating Disorder
WHO	World Health Organisation

CHAPTER ONE

Literature Review: Perinatal Mental Health and Screening

Although pregnancy has traditionally been perceived as a period of unconditional happiness and emotional wellbeing (Kendell, Wainwright, Hailey, & Shannon, 1976), there is accumulating evidence it does not offer protection from psychological distress and the development of mental health concerns (Bondas & Eriksson, 2001; Buist, 2014; Carin, Lundgren, & Bergbom, 2011; Giardinelli, Cecchelli, & Inncenti, 2008; Schneider, 2002; Melander & Lauri, 1999). Similar to their non-pregnant counterparts, pregnant women can experience an exacerbation, onset, or return of psychological distress and/or mental illnesses across the diagnostic spectrum (Giardinelli et al., 2008; Howard et al., 2014; Jones, Chandra, Dazzan, & Howard, 2014). Although mental health concerns are one of the most common morbidities during pregnancy and in the postnatal period (Howard et al., 2014; Jones et al., 2014), assessing maternal mental health and wellbeing during pregnancy has often been perceived as less important than ensuring optimal physical health for the mother and improving birth outcomes (Bauer, Parsonage, Knapp, Lemmi, & Adelaja, 2014; beyondblue, 2008; Davies, 2015; Hogg, 2013; Naylor et al., 2016). It is, however, well established that poor maternal mental health during the perinatal period, inclusive of pregnancy and the first-year post-birth, has negative effects and consequences for the mother, child, partner (if present), and immediate family (Bauer et al., 2015; beyondblue, 2008; Gavin et al., 2005; Gray, 2013; Lovestone & Kumar, 1993; Meltzer-Brody & Stuebe, 2014; Oates, 2015).

Poor mental health in the perinatal period is not only a significant concern for infants, parents, and families, but is also related to considerable financial burden for Australia and other countries (Bauer et al., 2014; beyondblue, 2008). Recent estimates indicate that untreated perinatal mental health conditions cost the Australian economy more than $538 million in 2013 alone (PricewaterhouseCoopers [PwC] & Centre of Perinatal Excellence

[COPE], 2014). Furthermore, when the ongoing impacts to mothers, children, and families were considered over a 20-year period, the cost of untreated perinatal mental health conditions increased to $710 million (PwC & COPE, 2014). It is predicted that a cost saving of $147 million over a two year period is possible if the prevalence of perinatal mental health concerns can be reduced by five percent using an early detection and intervention framework (PwC & COPE, 2014). As such, improving maternal mental health and wellbeing during pregnancy and the postnatal period has been a national priority in Australia for the past two decades, championed by advocates from the fields of mental health, midwifery, child and family health, general practice, and allied health community services (Austin, Highet, and the Guidelines Expert Advisory Committee, 2011; beyondblue, 2008).

Australia's focus on perinatal mental health is strongly linked to enacting and achieving two of the millennium development goals outlined by the United Nations (UN): improving and enhancing maternal health, and reducing child mortality (UN, 2014; World Health Organisation [WHO], 2009). A key barrier to supporting women in the perinatal period, however, is the poor rate at which mental health conditions are identified. It is estimated that up to 75 percent of women with a perinatal mental health condition are not identified (Coates et al., 2004; Spitzer et al., 2000) and only one in ten women who need mental health care receive it (Bowen et al., 2012). Acknowledging the urgent need for improved identification and treatment of mental health conditions in the perinatal period, the National Institute for Health and Care Excellence (NICE) recently published a quality standard pertaining to mental health pathways in pregnancy and postpartum (NICE, 2016). In this guideline, perinatal mental health was discerned as a significant priority for quality improvement, particularly screening and identification.

Importance of Identifying Perinatal Mental Health Conditions

Regardless of severity, maternal mental health difficulties have been shown to critically impact a woman, her child, and her family (Bauer et al., 2015; beyondblue, 2008; Gavin et al., 2005; Gray, 2013; Lovestone & Kumar, 1993; Meltzer-Brody & Stuebe, 2014; Oates, 2015). An emerging body of literature continues to highlight the importance and complexity of mental health in the perinatal period and the inter-connection between physical and mental health during this time (beyondblue, 2008; Davies, 2015; Hogg, 2013; Naylor et al., 2016). In addition to affecting the emotional wellbeing and happiness of mothers, mental health concerns during pregnancy affect the gestation experience (Austin, Highet, & the COPE Expert Working Group, 2017), are often associated with increased risk of undesirable obstetric and neonatal outcomes or complications (Gold & Marcus, 2008; Howard, Goss, Leese, & Thornicroft, 2003; Stein et al., 2014), and can profoundly affect the mother-infant attachment and the infant's cognitive and psychological development in the shorter and longer term (Barker et al., 2011; Laurent et al., 2013; Sinclair & Murray, 1998; Verbeek et al., 2012).

To support women, infants, and families to achieve optimal mental health and wellbeing in the perinatal period, both physical and psychological (including social) health must be considered in every aspect of maternity care (Austin et al., 2017). Due to the routine contact all women in Australia have with primary health care services during pregnancy, pregnancy and antenatal care represent a unique window of opportunity for promotion, prevention, and early intervention in mental health (Austin et al., 2017; beyondblue, 2008; NICE, 2017a). In Australia, primary health care in pregnancy is predominately delivered by general practitioners (GPs), midwives, maternal and child health nurses, and specialists such as obstetricians (Austin et al., 2011). Primary health care professionals also tend to act as the entry point to secondary mental health care services and health professionals such as

psychologists, psychiatrists, social workers, and mental health care nurses (Austin et al., 2011).

Prevalence of Mental Health Concerns in Pregnancy

Until recently, research and clinical care in perinatal mental health predominately focused on the postnatal period, with mental health problems during pregnancy being relatively neglected (NICE, 2017a). It is increasingly recognised, however, that pregnancy is not protective against the development of mental health conditions (NICE, 2017a), with large prospective community-based studies suggesting point prevalence rates for some conditions are higher during pregnancy than in the postnatal period (Evans, Heron, Francomb, Oke, & Golding, 2001; Heron et al., 2004).

Accurately specifying or estimating the overall prevalence of mental health conditions in pregnancy is difficult. For each condition, different prevalence rates are reported depending on the study design and sample size (e.g., large prospective population based studies vs. smaller longitudinal or cross-sectional studies), conceptualisation of the condition (e.g., anxiety could be conceptualised as stress, generalised anxiety, or fear of childbirth), the method of measurement (e.g., self-report inventories vs. clinical interviews), the clinical cut-off employed (particularly for screening or self-report instruments), and the timing of measurement (e.g., intrapartum vs. postpartum, which gestation week or trimester, or when during the postnatal period).

At the current time, it is broadly estimated that mental illness affects between 10 and 26 percent of women during the perinatal period (Bauer et al., 2014; Gavin et al., 2005; Vesga-López et al., 2008). This is not dissimilar to the 12-month prevalence rates estimated for mental health conditions in non-pregnant women in Australia. According to the 2007 National Survey of Mental Health and Wellbeing (Australian Bureau of Statistics, 2008), 22.3 percent of women experienced a mental disorder (of any type) in the previous 12

months, 17.9 percent experienced an anxiety disorder, 7.1 percent experienced a mood disorder, 5.1 percent experienced a substance use disorder, 1 to 2 percent experienced bipolar affective disorder, and less than 1 percent experienced a psychotic disorder. These statistics may, however, underestimate the prevalence of mental health conditions in Australia as data collection was more than a decade ago and significant efforts have been made to destigmatise and improve access to mental health services in Australia (Department of Health & Ageing, 2013). A 2012 report by Deloitte Access Economics, commissioned by the Butterfly Foundation, suggests the prevalence of eating disorders (EDs) in the general population is around 4 percent at any given time, with a lifetime prevalence of 9 percent (Butterfly Foundation, 2012; Wade, Crosby, & Martin, 2006). The prevalence of subclinical disordered eating is thought to be much higher, affecting up to 20 percent of women (Wade, Bergin, Tiggemann, Bulik, & Fairburn, 2006; Koupil, Tooth, Heshmati, & Mishra, 2016).

When individual conditions during pregnancy are examined, depression has the largest evidence base. It is estimated that approximately 9 percent of Australian women experience depression during pregnancy (Australian Department of Health & Ageing, 2012; Buist & Bilszta, 2006), with this rate ranging from 4 to 20 percent internationally (see Biaggi, Conroy, Pawlby, & Pariante, 2016, for a full review). Anxiety disorders are also prevalent, with estimates ranging from 5 to 21 percent during pregnancy (beyondblue, 2008; Giardinelli et al., 2012; Heron et al., 2004). Comorbidity is also common during pregnancy, with two-thirds of women with depression during pregnancy also having an anxiety disorder (Lydsdottir et al., 2014; Wisner et al., 2013). While the prevalence of other serious mental illnesses such as bipolar disorder, schizophrenia, and borderline personality disorder are much less common (approximately 1%; Galletly et al., 2016; Mitchell et al., 2013) than depression and anxiety disorders during pregnancy, these conditions have as much, if not more, of a negative impact on maternal and infant outcomes, particularly when a woman has

serious or multiple adverse psychosocial circumstances (see Rusner, Berg, & Begley, 2016, for a review).

Similarly, EDs represent significant mental health concerns that often afflict women during prime childbearing years (Abebe, Lien, & von Soest, 2012; Hsu, 1989; Leddy, Jones, Morgan, & Schulkin, 2009; Stice, Marti, & Rohde, 2013). Consequently, the expression of disordered eating symptomatology is possible during pregnancy, whether symptoms were existing, exacerbated, previously active, or only emerged during this period (Tierney, Fox, Butterfield, Stringer, & Furber, 2011). While scientific literature examining the intersection between EDs and pregnancy has grown over the past two decades, overall, there is a paucity of research in this area (Watson et al., 2014). Furthermore, clinical recommendations and resources for practitioners to identify and support women presenting with disordered eating symptomatology in antenatal care are limited. Consistent with the priorities of the UN (2014), WHO (2009), and NICE (2016), the overarching aim of this thesis was to improve the screening and identification of disordered eating in pregnancy.

Prevalence of Eating Disorders in Pregnancy

The term "eating disorder" refers to a range of conditions that often entail profound physical, psychological, and social effects (Agras, 2001). In a broad sense, EDs can be described as unhealthy attitudes and disturbances toward eating, shape, and weight, in addition to distorted body image (American Psychiatric Association [APA], 2013; Watson et al., 2014). This often results in significant departures from normal eating behaviour (e.g., dietary restriction, binge eating, and purging) and other compensatory weight control behaviours such as excessive exercise, and misuse of laxatives, diuretics, or enemas (APA, 2013; Easter, 2015).

The fifth edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM-5; APA, 2013) recognises three primary EDs: anorexia nervosa (AN), bulimia nervosa

(BN), and binge eating disorder (BED). Presentations that do not fit within these diagnoses (approximately 20% to 40% of cases) are classified under residual categories known as other specified or unspecified feeding or eating disorder (OSFED or USFED; Culbert, Racine, & Klump, 2015). In previous editions of the DSM, namely the fourth edition (DSM-IV-TR), OSFED was conceptualised as eating disorder not otherwise specified (EDNOS). Although diagnostic distinctions between these conditions are made, a transdiagnostic overlap between conditions is often observed (Fairburn, Cooper, & Shafran, 2003). That is, several component symptoms (e.g., overvaluation of weight/shape, body dissatisfaction, dietary restriction, binge eating) are often shared across diagnoses. An overview of the diagnostic criteria for each condition is shown in Table 1.

Until recently, the relationship between EDs and pregnancy was poorly understood and lacking in empirical literature; however, crucial scientific insight has emerged from two prospective population-based mother and child cohort studies: the Norwegian Mother Baby (MoBa) cohort (N > 100,000) which commenced data collection in 1999 and currently collects data from 50 of the 52 hospitals in Norway (Magnus et al., 2006; Watson et al., 2012), and the Avon Longitudinal Study of Parents and Children (ALSPAC) in the United Kingdom (N = 14,541), which commenced data collection in 1991 and is ongoing to date (Golding, Pembrey, & Hones, 2001). With literature and understanding in this area growing, ED symptoms during pregnancy have been linked to a range of negative consequences such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and other obstetric and postpartum difficulties (Linna et al., 2014; Watson et al., 2014).

Table 1

Summary of the DSM-5 Diagnostic Criteria for Eating Disorders Including Clinical Features and Sex Ratios

	AN	BN	BED	OSFED
Diagnostic criteria (DSM-5)	• Persistent energy restriction resulting in significantly low body weight (weight that is less than minimally normal, or minimally expected) • Intense fear of gaining weight or of becoming fat, or persistent behaviour that interferes with weight gain • Cognitive distortions related to body shape and weight perception, undue influence of body weight and shape on self-perception, or persistent lack of recognition of the seriousness of current low body weight	• Recurrent episodes of binge eating, characterised by consumption of an objectively large amount of food within a discrete period of time and a sense of lack of during the episode • Recurrent inappropriate compensatory behaviours to prevent weight gain (e.g., self-induced vomiting; misuse of laxative, diuretics, or enemas; excessive exercise; or fasting) • Binge eating and compensatory behaviours occur at least once a week for three months • Self-evaluation is unduly influenced by body shape and weight	• Recurrent episodes of binge eating, characterised by at least three of the following: ○ Eating more rapidly than usual ○ Eating until uncomfortably full ○ Eating large amounts of food when not physically hungry ○ Eating alone because of embarrassment ○ Feeling disgusted, depressed, or very guilty afterwards • Marked distress regarding binge eating • Binge eating occurs at least once a week for three months • Binge eating is not associated with compensatory behaviours	• Presentations in which symptoms are characteristic of feeding or eating disorders, but do not meet the full diagnostic criteria for any specific ED • Symptoms cause clinically significant distress • Can include: ○ Atypical AN ○ BN (of low frequency and/or limited duration) ○ BED (of low frequency and/or limited duration) ○ Purging disorder ○ Night eating syndrome
Clinical presentation	• Low weight status • Amenorrhea may or may not be present • Medical comorbidities	• Normal body weight • Amenorrhea may or may not be present • Sometimes potassium depletion and dental caries	• Typically overweight or obese, but can also be normal weight	• Varies from case to case
Sex ratio (female: male)	10:1	10:1	3:2	Unknown

While exploring a range of health outcomes, data from the MoBa cohort revealed the prevalence of an ED during pregnancy was approximately 5.1 percent (4.8% BED, 0.2% BN, and 0.1% purging disorder [EDNOS-P]) when using non-standardised broad diagnostic criteria aligned to the fourth edition of the DSM (Bulik et al., 2007). The prevalence of AN was not measured due to diagnostic difficulties associated with the DSM-IV weight criterion during pregnancy. Notably, these rates were also similar to the prevalence of EDs in the general population in Norway at the time (Götestam & Agras, 1995), indicating pregnancy does not always function as a reprieve from disordered eating concerns or further exacerbate such symptoms. A subsequent MoBa study incorporating new data cases since the initial publication has supported the documented prevalence rates of EDs in the MoBa cohort (Watson et al., 2012), demonstrating the ongoing pattern of disordered eating in pregnant women in Norway.

Around the same period as the initial MoBa prevalence study, data from the ALSPAC cohort in the United Kingdom (UK) revealed a prevalence rate of 0.05 percent for AN and 0.4 percent for BN at 12 weeks gestation (Micali, Treasure, & Simonoff, 2007). While the prevalence of BN was consistent with the findings of Bulik et al. (2007), reliance on self-reported DSM-IV threshold ED diagnoses in Micali et al. (2007), both recent and historical episodes, is likely to have resulted in an underestimation of EDs during the antenatal period, particularly women who have not received a formal diagnosis and/or those with subthreshold symptoms. Other issues in Micali et al. (2007) included the absence of prevalence rates for BED or other/unspecified EDs, which did not appear to be investigated, and a lack of clarity around the operationalisation of ‘recent ED history’. That is, it was not explicitly stated whether the ED assessment period related to pre-pregnancy or the first three months of the pregnancy.

More recently, a smaller cross-sectional study undertaken in a prenatal clinic in the UK screened women for an ED at the first routine antenatal appointment in the first trimester using the Eating Disorder Diagnostic Scale (EDDS; Stice, Fisher, & Martinez, 2004), a self-report instrument that aligns to the diagnostic criteria of the DSM-IV. Of the 739 women screened, 7.5 percent (1 in 14) were found to meet criteria for an ED (Easter et al., 2013). Specifically, 5 percent were diagnosed with EDNOS (mostly subclinical BED), 1.8 percent with BED, 0.5 percent with AN, 0.1 percent with BN, and 0.1 percent with purging disorder (Easter et al., 2013). Although the overall prevalence estimate was higher due to the inclusion of EDNOS presentations, the estimates for specific ED subtypes are similar to the MoBa and ALSPAC samples (Watson et al., 2014), indicating further research exploring the identification, assessment, management, and prevention of EDs and subclinical variants in pregnancy is warranted. However, similar to Bulik et al. (2007) and Micali et al. (2007), issues with self-report methodologies, particularly use of generic ED instruments designed for and validated in non-pregnant populations, likely resulted in an underestimation of the true population prevalence of EDs in pregnancy in all three studies. Furthermore, these prevalence rates do not necessarily reflect rates of subthreshold disordered eating in pregnancy, which may also result in negative neonatal and maternal outcomes.

Operationalisation of Disordered Eating in Pregnancy

The diagnostic disorder approach to mental health aims to classify disorders into distinct and separate categories (Goldberg, 2010). The intention behind this approach is to facilitate ongoing revision of the diagnostic criteria for each condition when new research surfaces (Coates, 2017). Theoretically, this is meant to ensure that individuals who meet criteria for a specific condition are identified, providing estimates of prevalence, incidence, and prognosis, and encouraging intervention (Goldberg, 2010). One of the main issues with the disorder classification approach and therefore prevalence data is that conditions are

dichotomised as being present or absent when, in reality, an individual's symptoms may occur along a continuum (diFlorio & Meltzer-Brody, 2015).

Contemporary theorists argue that EDs represent a spectrum of symptoms that fall along a continuum with healthy beliefs, attitudes, and behaviours toward body weight and shape, eating, and exercise-related content at one end of the spectrum and problematic beliefs, attitudes, and behaviours (and ultimately syndromes) at the opposite end (Alexander & Treasure, 2012; National Eating Disorders Collaboration [NEDC], 2016). Along this continuum is subthreshold disordered eating, which has also been described as subclinical maladaptive eating (Chamay-Weber et al., 2005; Shisslak, Cargo, & Estes, 1995). Disordered eating has typically been defined as a wide range of eating behaviours and cognitions resulting in a negative impact on an individual's emotional, social, and physical wellbeing (APA, 2013). Typically, the distinction between disordered eating and a threshold ED has been the degree of severity, with disordered eating symptoms occurring at a lower level of severity and/or frequency (Austin, 2000; Butterfly Foundation, 2017; Mustapic, Marcinko, & Vargek, 2015; Stice, Shaw, & Marti, 2007; Watkins & Lask, 2002).

Disordered eating can also represent changes in eating and exercise patterns due to developmental stages (e.g., pregnancy, early childhood, and advancing age), other mental health conditions (e.g., major depressive disorder), or certain life events (e.g., moving away from home, relationship breakdown). In these circumstances, the changes in an individual's eating and/or exercise patterns are typically transient and/or not accompanied by significant psychological or physical distress (NEDC, 2016). In relation to pregnancy, most women report disturbances in normal eating patterns (Tiller & Treasure, 1998), usually in the form of food cravings, increases or decreases in appetite, changes to dietary preferences, inconsistent eating patterns, food aversions, and nausea and vomiting (Dickens & Trethowan, 1971; Fairburn et al., 1992). As shown in Table 2, such symptoms are also observed in the context

of disordered eating (Easter et al., 2013; Easter, 2015; Watson et al., 2014), resulting in a clinical overlap of symptoms. These disturbances are often, however, considered 'normal' within the context of pregnancy due to hormonal fluctuations, changes in sensory perception, and maternal and/or fetal nutritional needs (see Orloff & Hormes, 2014, for a review), making it difficult to delineate normative pregnancy changes from disordered eating symptomatology.

Table 2

Clinical Overlap Between Disordered Eating and Pregnancy Symptoms

Disordered Eating Symptoms	Pregnancy-Related Symptoms
• Purging behaviours (e.g., self-induced vomiting)	• Nausea and vomiting (sometimes self-induced)
• Dietary restriction	• Decreased appetite
• Food aversions and/or avoidance of particular foods	• Food aversions and/or avoidance of particular foods
• Binge eating behaviours	• Increased appetite / "eating for two"
• Cognitions dominated by food- or body-related content	• Cognitions dominated by food- or body-related content
• Emotion fluctuations	• Emotion fluctuations
• Lethargy	• Lethargy

The typical onset and duration of eating and food-related disturbances that often occur during pregnancy has been reported in some studies (see Table 3) and may represent a possible avenue to distinguish normative pregnancy experiences from disordered eating symptomatology. While these studies provide some indication of the gestational onset and duration of such changes, the temporal pattern of certain symptoms does differ between studies, often considerably. As such, clinical identification of disordered eating based on deviations outside these temporal ranges is problematic. Additionally, it remains unclear whether pregnancy-appropriate 'abnormal' eating behaviours that fall outside the normal

temporal pattern would immediately constitute a "disordered" or "problematic" label as these overt behaviours alone would not encapsulate the complexity of disordered eating symptomatology, which often encompasses a range of behavioural, cognitive, and affective components.

Table 3

Temporal Pattern of Frequently Reported Pregnancy Symptoms

Pregnancy Symptom	Prevalence (%)	Onset period (Wk of gestation)	Reference/s
Food cravings	50-54	9 to 12	Abraham et al. (1994)
		6 to 28	Bayley, Dye, Jones, DeBono, & Hill (2002); Pope et al. (2002); Belzer et al. (2010)
		9 to 19.5	Fairburn, Stein, & Jones (1992)
Food aversions	80	7 to 18.5	Fairburn et al. (1992); Crystal, Bowen, & Berstein (1999)
Nausea	60-90	6 to 16	Abraham et al. (1994); Bayley et al. (2002); Fairburn et al. (1992); Tierson et al. (1986)
Nausea and vomiting	50-56	6 to 10	Bayley et al. (2002); Stein & Fairburn (1996)
		6 to 20	Fairburn et al. (1992); Hawkins & Gottlieb (2013)

Overall, there continues to be a lack of clarity and guidance as to how clinicians and researchers distinguish pregnancy-appropriate symptomatology from eating disordered-related thoughts, attitudes, and behaviors in pregnancy populations. Such uncertainty and lack of operationalisation could lead to poor recognition and management of EDs, particularly subthreshold syndromes, which some researchers have suggested negatively impact pregnancy outcomes to a similar extent as full threshold syndromes (Crow et al., 2008; Harris, 2010).

Attempts to clarify disordered eating behaviours in pregnancy have also led to the emergence of misleading terms such as "pregorexia". Pregorexia describes women who have an excessive fear of pregnancy-related weight gain and use various methods (e.g., extreme

exercise regimes and dietary restriction) to avoid or disrupt weight increases that characterise the course of a normal, healthy pregnancy (Hall-Flavin, 2015; Mathieu, 2009; Wallace, 2013). While the term has circulated in popular media over the past decade, it is not a formally recognised ED in the DSM-5 and typically describes women with features characteristic of AN and BN, thereby failing to consider binge eating behaviours, the most prevalent eating disturbance during pregnancy (Bulik et al., 2007; Knoph Berg et al., 2011). The terminology may also stigmatise and trivialise a complex condition/experience; and is therefore unlikely to facilitate symptom disclosure from women in antenatal settings.

Prevalence of Disordered Eating in Pregnancy

Despite this conceptual ambiguity, the prevalence of disordered eating behaviours and cognitions has been assessed in a small number of studies over the past two decades, predominately via administration of self-report psychometric instruments. Similar to other mental health conditions in pregnancy, these studies have revealed somewhat conflicting evidence depending on the characteristics of the sample (i.e., pregnancy stage), component of disordered eating being investigated (e.g., cognitive vs. affective), the psychometric instrument employed (e.g., screening tool vs. self-report inventory vs. clinical interview), and the various instrument thresholds used to determine clinically significant scores. Due to these methodological discrepancies, comparison between studies is challenging; however, estimates of disordered eating in pregnancy using global measures range from 0.6 percent or less than one in 20 pregnant women, to 27.8 percent or 5.5 in every 20 pregnant women (Broussard, 2012; Easter et al., 2013; Fairburn, Stein, & Jones, 1992; Micali et al., 2007; Pettersson, Zandian, & Clinton, 2016; Soares et al., 2009; Turton, et al., 1999). A summary of these prevalence studies is detailed in Table 4.

Table 4

Summary of Studies Exploring the Prevalence of Disordered Eating in Pregnancy

Study	Methodology	Findings / Prevalence Estimates

Fairburn, Stein, & Jones (1992)	• Prospective – two assessment time points
	  ○ 15 weeks and 32 weeks gestation
	• Instrument – modified version of the EDE-Q
	  ○ Clinical cut-off score used not stated
	• Sample – 100 pregnant women (UK)
		• 5.0% experienced objective binge/purge episodes
		• 3.0% reported self-induced vomiting during pregnancy
		• 0.9% reported clinically significant dietary restraint
		• 1.1% reported clinically significant shape concerns
		• 0.8% reported clinically significant weight concerns

Turton, Hughes, Bolton, & Sedgwick (1999)	• Observational cross-sectional – one assessment time point
	  ○ Mean gestational week of administration not reported
	• Instrument – EAT
	  ○ Clinical cut-off score of 19 used
	• Sample – 530 pregnant women (UK)
		• 4.9% scored above the clinical threshold on the EAT, indicative of clinically significant disordered eating concerns.

Kelly, Zatzick, & Anders (2001)	• Observational cross-sectional – one assessment time point
	  ○ Mean gestational week of administration not reported
	• Instrument – PRIME-MD Patient Health Questionnaire
	  ○ Only measures symptoms consistent with bulimia nervosa and binge eating disorder
	• Sample – 186 pregnant women (US)
		• 6.0% displayed disordered eating symptoms consistent with bulimia nervosa (2%) and binge eating disorder (4%)

Note. EDE-Q = Eating Disorders Examination Questionnaire, EAT = Eating Attitudes Test, EDI-3 = Eating Disorders Inventory-3, EDDS= Eating Disorders Diagnostic Scale.

Table 4 (continued)

Summary of Studies Exploring the Prevalence of Disordered Eating in Pregnancy

Study	Methodology	Findings / Prevalence Estimates

Micali, Treasure, & Simonoff (2007)	• Prospective cohort study – two assessment time points
	  ○ 18 weeks and 32 weeks gestation
	• Instruments – various
	  ○ Weight & Shape subscales of the EDE-Q (18 weeks)
	  ▪ EDE-Q cut-off of 2 used
	  ○ Study specific self-report screening items (18 & 32 weeks)
	• Sample – 12,252 pregnant women (UK)
		• 2.4% reported laxative use during pregnancy
		• 7.9% reported self-induced vomiting during pregnancy
		• 26.2% were classified as 'high exercisers'
		• 46.3% reported feeling a loss of control over eating
		• 65.5% reported dieting during pregnancy
		• 38.6% had a strong desire to lose weight while pregnant
		• 65.2% perceived they had put on "too much" weight during pregnancy
		• 67.0% were anxious about gaining weight during pregnancy

Soares et al. (2009)	• Prospective cohort study, however, measurement of disordered eating was cross-sectional – one assessment time point
	  ○ 16 to 32 weeks gestation
	• Instrument – modified version of the EDE-Q
	  ○ Clinical cut-off score of 4 used
	• Sample – 712 pregnant women (Southern Brazil)
		• 17.3% experienced objective binge eating episodes
		• 1.8% reported self-induced vomiting
		• 0.1% misused laxatives
		• 0.0% misused diuretics
		• 1.3% had clinically significant dietary restraint scores
		• 0.1% had clinically significant eating concerns
		• 5.6% had clinically significant shape concerns
		• 5.5% had clinically significant weight concerns.
		• 0.6% were considered to have a possible threshold ED

Note. EDE-Q = Eating Disorders Examination Questionnaire, EAT = Eating Attitudes Test, EDI-3 = Eating Disorders Inventory-3, EDDS= Eating Disorders Diagnostic Scale.

Table 4 (continued)

Summary of Studies Exploring the Prevalence of Disordered Eating in Pregnancy

Study	Methodology	Findings / Prevalence Estimates

Broussard (2012)	• Retrospective – one assessment time point
	  ○ Asked to reflect on pre-pregnancy and pregnancy experience
	• Instrument – EDI-3
	  ○ Clinical cut-off scores: i) Bulimia ≥ 5, ii) Body dissatisfaction ≥ 22, iii) Drive for thinness ≥ 17
	  ○ Subscale elevation indicative of disordered eating
	• Sample – 54 postpartum women (US)
		• 27.8% of the sample had EDI-3 scores indicating the presence of psychological and behavioural traits associated with EDs during pregnancy.

Easter et al. (2013)	• Observational cross-sectional – one assessment time point
	  ○ 10 to 12 weeks gestation (*M* = 11.5 weeks)
	• Instrument – modified version of the EDDS
	  ○ Clinical cut-off score of 4 used
	• Sample – 739 pregnant women (UK)
		• 5.0% met criteria for eating disorder not otherwise specified
		• 8.4% reported binge eating with loss of control (≥ 2/week)
		• 0.9% reported fasting to counteract the effects of eating or to prevent weight gain
		• 0.7% reported excessive exercise to counteract the effects of eating or to prevent weight gain
		• 1.1% reported self-induced vomiting to counteract the effects of eating or to prevent weight gain
		• 0.3% reported use of laxatives or diuretics to counteract the effects of eating or to prevent weight gain
		• 23.4% scored above the clinical cut-off for high weight and shape concern

Note. EDE-Q = Eating Disorders Examination Questionnaire, EAT = Eating Attitudes Test, EDI-3 = Eating Disorders Inventory-3, EDDS= Eating Disorders Diagnostic Scale.

Table 4 (continued)

Summary of Studies Exploring the Prevalence of Disordered Eating in Pregnancy

Study	Methodology	Findings / Prevalence Estimates

Bvakhbakhi, Liebling, & Morgan (2014)	• Observational cross-sectional – one assessment time point
	  ○ Mean gestational week of administration not reported
	• Instrument – Whooley questions
	  ○ Identification of disordered eating not detailed
	• Sample – 158 pregnant women (UK) identified 'at risk' during routine perinatal screening
		• 6.0% of the at risk notifications related to disordered eating.

Mohamadirizi et al. (2015)	• Observational cross-sectional – one assessment time point
	  ○ Mean gestational week of administration not reported
	• Instrument – EDE-Q
	  ○ Cut-off point used in study not stated
	• Sample – 213 pregnant women (Iran)
		• 8.5% of the sample displayed disordered eating symptoms

Pettersson, Zandian, & Clinton (2016)	• Observational cross-sectional – one assessment time point
	  ○ 10 to 12 weeks gestation
	• Instrument – EDE-Q (traditional & optimised versions)
	  ○ Clinical cut-off score of 2.8 used
	• Sample – 396 pregnant women (Sweden)
		• Using the traditional 22-item EDE-Q, 3.0% of the sample was suspected to have clinically significant disordered eating concerns
		• Using the optimised 14-item EDE-Q, 5.3% of the sample was suspected to have clinically significant disordered eating concerns

Note. EDE-Q = Eating Disorders Examination Questionnaire, EAT = Eating Attitudes Test, EDI-3 = Eating Disorders Inventory-3, EDDS= Eating Disorders Diagnostic Scale.

Overall, these studies demonstrate that disordered eating, when conceptualised as subclinical ED symptoms, is relatively common during pregnancy. However, given that each study utilised psychometric instruments designed for and validated in non-pregnant populations, it is plausible these estimates are an under- or over- representation of the true prevalence of disordered eating symptomatology in pregnancy. Robust and accurate investigation of the prevalence of particular conditions or phenomena is well known to be inherently challenging, particularly when there is a high level of heterogeneity in the phenomena of interest and/or the scientific methodologies employed across studies (Fletcher, 2007; Seck et al., 2016; Ward, 2013). To unify research efforts in this area, standardised psychometric instruments designed to assess/measure disordered eating in pregnancy must be employed. This would also improve epidemiological accuracy through enhanced measurement validity, in addition to facilitating measurement consistency across studies.

Thus, to date, there appears to be no clear definition or operationalisation of disordered eating in pregnancy. Clarifying this distinction is important to improve epidemiological accuracy in research, particularly if a psychometrically sound pregnancy-specific assessment measure is to be developed, in addition to facilitating comparison of literature and findings across research and clinical settings.

Course of Disordered Eating and EDs During the Perinatal Period

Intrapartum. To date, literature regarding the course of disordered eating symptoms during pregnancy (intrapartum) has been conflicting. Some studies suggest pregnancy may improve disordered eating symptoms and potentially serve as a catalyst for remission (Blais et al., 2000; Bulik et al., 2007; Crow et al., 2008; Crow, Keel, Thuras, & Mitchell, 2004; Micali et al., 2007; Rocco et al., 2005; Soares et al., 2009). Other studies report pregnancy may represent a period of increased vulnerability, potentially exacerbating disordered eating symptomatology for some women (Benton-Hardy & Lock, 1998; Coker, Mitchell-Wong, &

Abraham, 2013; Conrad, Schablewski, Schilling, & Liedtke, 2003; Senior et al., 2005; Tiller & Treasure, 1998). Regardless of these conflicting findings, it has been suggested that women with a pre-existing ED continue to experience an increased level of ED psychopathology during pregnancy and a complete absence of ED symptomatology for these women is rare (Blais et al., 2000; Crow et al., 2008; Micali, 2010; Tierney et al., 2013).

While some of the behavioural components of an ED (e.g., compensatory behaviours) have been shown to improve, high levels of weight and shape concern often persist during pregnancy (Blais et al., 2000; Crow et al., 2008; Easter et al., 2013; Micali, 2010; Micali et al., 2007). Such findings were demonstrated in Easter et al. (2013), with compensatory behaviours such as fasting, excessive exercise, and the misuse of laxatives or diuretics, typically remitting or reducing during pregnancy, compared to self-reported behaviours engaged in three months prior to pregnancy. Notably, this pattern was not observed for self-induced vomiting and binge eating behaviours, both of which increased during pregnancy. Possibly endocrinological factors associated with pregnancy trigger nausea, vomiting, and increased appetite, thereby increasing the difficulty of extinguishing ingrained, pathological versions of these behaviours (Bulik et al., 2007).

Easter et al. (2013) also revealed that ED-related cognitions were slightly higher during pregnancy, with 173 women (23.4%) scoring above the cut-off for high weight and shape concern during the intrapartum period, whereas 159 women (21.5%) had scored above the cut-offs in the three months prepartum. This is not unexpected given Easter et al. (2013) collected data during the first trimester, which research suggests coincides with increased body dissatisfaction due to difficulties associated with the "in-between" stage (i.e., the absence of a visible "baby bump"; Darvill et al., 2010). That is, the stomach swells and women fear others might misinterpret this as "fat-related" weight gain. Following this in-between stage, women's body satisfaction appears to improve and stabilise in the second

trimester and the first half of the third trimester, and then declines again toward the end of pregnancy as women worry and ruminate about postpartum weight loss and returning to their pre-pregnancy body state (Clark & Ogden, 1999; Darvill et al., 2010; Goodwin, Astbury, & McMeeken, 2007; Rocco et al., 2005; Skouteris, Carr, Wertheim, Paxton, & Duncombe, 2005). However, this is not a consistent trajectory for all women and is dependent on a range of factors including pre-pregnancy body mass index (Duncombe, Wertheim, Skouteris, Paxton, & Kelly, 2008; Fox & Yamaguchi, 1997; Haedt & Keel, 2007; Micali et al., 2007), history of disordered eating (Baker, Carter, Cohen, & Brownell, 1999; Fairburn & Welch, 1990), psychological wellbeing (Clark et al., 2009; DiPietro et al., 2003; Duncombe et al., 2008; Haedt & Keel, 2007; Skouteris et al., 2005), degree of social support (DiPietro et al., 2003; Lai et al., 2005), and level of physical activity (Boscaglia, Skouteris, & Wertheim, 2003; Goodwin et al., 2000).

Postpartum. Research has indicated that although antenatal improvements in ED psychopathology may be retained postnatally for some women (Blais et al., 2000; Lemberg & Phillips, 1989), relapse to pre-pregnancy ED levels or worsening of disordered eating symptoms is common following birth, particularly within the first six to 12 months (Astrachan-Fletcher, Veldhuis, Lively, Fowler, & Marcks, 2008; Micali et al., 2007). Furthermore, an ED or disordered eating symptomatology may develop in the postnatal period (NEDC, 2015). Notably, research suggests the prevalence of disordered eating in the postpartum period ranges from 7.8 to 12.8 percent in general pregnancy populations (i.e., individuals without a lifetime history of disordered eating; Larsson & Andersson-Ellström, 2003; Pettersson et al., 2016). For women with a lifetime history of an ED or disordered eating concerns, relapse or resurgence of symptoms during the postpartum is reportedly as high as 57 percent (Blais et al., 2000; Lacey & Smith, 1999; Morgan, Lacey, Sedgwick, 1999). In addition to the rapid hormonal changes and psychosocial stressors that every

mother encounters immediately postpartum, which may increase the risk of a developing psychological problems, the absence of having the child 'within' the mother, and thus the desire to "protect" the unborn child from the ED, may partially account for the relapse of many women (Astrachan-Fletcher et al., 2008). That is, the mother no longer considers her body to be serving a functional, life-sustaining role for the valued life of her child and reverts to an evaluative appraisal of herself and her body, characterised by undervaluation and punishment (Mason, Cooper, & Turner, 2012).

Specific trajectories from prepartum to postpartum. Findings from the MoBa cohort have also provided insight into the specific trajectory of EDs from prepartum, intrapartum, and into postpartum using a prospective longitudinal study design (Bulik et al., 2007; Knoph et al., 2013; Watson et al., 2012). For example, Bulik et al. (2007) revealed prepartum prevalence estimates in the MoBa cohort ranged from 0.1 percent for AN, 0.7 percent for BN, 3.5 percent for BED, and 0.1 percent for EDNOS-P. The authors note the lower prevalence of prepartum AN in the sample (compared to previous epidemiological studies) may be related to issues of infertility women with AN often experience, which can decrease the likelihood of pregnancy and thus reduce prevalence statistics in these cohorts (Stewart et al., 1990; Strimling, 1984; Weinfeld, Dubay, Burchell, Mellerick, & Kennedy, 1977). During pregnancy (intrapartum), these estimates reduced for BN (0.2%) and EDNOS-P (< 0.1%), with the most common pattern being full or partial remission of the conditions during pregnancy, rather than continuation of symptoms. This potentially suggests pregnancy is a powerful stimulus to extinguish or reduce bulimic symptomatology for some women, but not all.

Unlike BN and EDNOS-P, the most common course for BED in Bulik et al. (2007) was continuation of symptoms rather than remission. This suggests pregnancy may represent a period of vulnerability or high-risk for some women, particularly given the range of

adaptive neuroendocrinological changes that occur during pregnancy (Orloff & Hormes, 2014). As noted earlier, AN was not assessed during pregnancy in Bulik et al. (2007) due to issues with accurate assessment of the DSM-IV weight criterion during the pregnancy period. This is a limitation of previous, and potentially current, ED diagnostic criteria for pregnancy populations. It highlights a need for diagnostic guidance and/or special provisions when assessing women who present with disordered eating symptoms in pregnancy.

Pregnancy as a High-Risk Period for Unaffected Women

While most women affected by a threshold ED during pregnancy are likely to have experienced a threshold ED before conceiving, pregnancy can also serve as a risk factor for the onset of disordered eating symptomatology (Fairburn, Welch, Doll, Davies, & O'Connor, 1997; Nunes, Pinheiro, Hoffman, & Schmidt, 2014; Tiller & Treasure, 1998). For example, Bulik et al. (2007) revealed incident cases of BED and binge eating symptoms in the MoBa cohort were much more common than the development of other EDs during pregnancy (approximately 1.12 new cases per 1000 person-weeks). Utilising the same prospective pregnancy cohort, Knoph Berg et al. (2011) revealed 4.1 percent *(n* = 1887) of the MoBa sample met criteria for broadly defined BED, with 49.3 percent of this 4.1 percent (*n* = 931) considered to be incident cases of BED during pregnancy. Tiller and Treasure (1998) also reported that pregnancy precipitated an ED in 23 patients examined, most of whom were diagnosed with restrictive type AN. This highlights the onset of an ED or subclinical disordered eating can occur at any stage during pregnancy. It also emphasises the importance of routine screening for disordered eating throughout the antenatal process for all women, with or without a history of an ED, including screening for higher weight spectrum conditions such a binge eating disorder.

Consequences of ED Symptoms During Pregnancy

For pregnant women suffering from an ED or subclinical disordered eating concerns, there are several increased risks, physically and psychologically, to the mother and unborn child across the perinatal period, including the conception process.

Pre-pregnancy outcomes. Research utilising the Norwegian MoBa cohort has indicated women with an ED, particularly AN, carry a two-fold increase of experiencing an unintended pregnancy, compared to unaffected women (Bulik et al., 2010). It has been suggested the increased risk of unintended pregnancies in EDs, particularly low weight spectrum conditions, may be due to patients (and potentially health professionals) incorrectly assuming conception cannot occur in the absence of menstruation (or irregular menstruation), leading to lower or less stringent contraception use (Bulik et al., 2010). This is concerning given unintended pregnancies have been associated with a number of maternal and infant risks including higher odds of maternal mental illness, maternal and child mortality, impaired child growth, child abuse, lack of and early cessation of breastfeeding, and delayed and inadequate prenatal care (Bahk et al., 2015; Gipson, Koenig, & Hindin, 2008; Klima, 1998; Sawyer, Tully, & Collin, 2001), before the potential impact of an ED on the pregnancy process has been considered. A recent retrospective population-based cohort study in the US (N = 743,630) has also revealed that both low and high pre-pregnancy BMI were associated with a small increased risk of severe maternal illness and death (see Lisonkova et al., 2017), with BMI often influenced by disordered eating symptomatology.

Pregnancy outcomes. Although literature in the area is varied and often discordant, current research suggests that threshold EDs affect birth outcomes, including increasing the risk of miscarriage, premature labour, and obstetric complications (Bulik et al., 1999; Bulik et al., 2009; Micali et al., 2007). While it is difficult to determine the unique impact of each ED, there is some evidence to suggest specific EDs affect birth outcomes differently. For

example, both AN and BN have been associated with the birthing of babies that are significantly lower in body weight (Linna et al., 2014; Micali et al., 2007; Solmi, Hatch, Hotopf, Treasure, & Micali, 2013), while BED has been associated with higher weight babies (Linna et al., 2014). Maternal AN has also been associated with intrauterine growth restriction, slow and poor fetal growth, premature contractions, premature birth, and perinatal mortality (Linna et al., 2014). Maternal BN has been associated with premature contractions, newborn resuscitations, birth defects, and low APGAR scores (Linna et al., 2014). Collectively, literature suggests women with all EDs during pregnancy are at increased risk of experiencing undesirable birth and obstetric outcomes. While it is unknown how subclinical disordered eating may affect birth and obstetric outcomes, researchers have postulated similar outcomes to threshold EDs are likely (Crow et al., 2008; Harris, 2010).

In addition to the impact on birth outcomes, women experiencing disordered eating and threshold EDs during pregnancy commonly experience comorbid depression and anxiety. Carter, McIntosh, Joyce, Frampton, and Bulik (2003) revealed that 40 percent of women with a threshold ED experienced a major depressive episode during the perinatal period. Similar findings were also revealed in Mazzeo et al. (2006), with 39 and 59 percent of women with AN and BN, respectively, experiencing antenatal depression. In larger epidemiological samples, Micali et al. (2010) and Easter et al. (2015) have shown that women experiencing disordered eating symptomatology during pregnancy had increased levels of depression and anxiety during and following pregnancy, with this risk exacerbated for women with a lifetime history of depression. As elevated stress and anxiety during pregnancy have been linked to poor birth outcomes and long-term infant health and developmental complications, it is important that antenatal practitioners are aware of and manage such concerns (see Davis & Sandman, 2010, for a review). In understanding the link between disordered eating and comorbid affective symptoms, it is important that antenatal practitioners actively screen for

disordered eating when a woman presents with anxious or depressive symptomatology during pregnancy (Andersen & Ryan, 2009; Hawkins & Gottlieb, 2013; Knoph-Berg et al., 2011; NICE, 2017b; NEDC, 2015; Ward, 2008)

Postpartum outcomes. On top of the risks that disordered eating carries during the intra-partum period, there are a number of additional risks affected women potentially face following the birth of their child. Given the elevated prevalence of co-morbid mood disorders among ED populations, it has been shown that women who experience an ED or subclinical disordered eating during pregnancy are at an increased risk of developing postnatal depression and, as a result, may struggle to adjust to their new role as mother. In Morgan et al. (1999), 33 percent of women with EDs during pregnancy experienced postnatal depression. Similar findings were also revealed by Franko et al. (2001), with 50 percent of women with active disordered eating during and after pregnancy found to develop postnatal depression, compared to 29 percent of women with a history of an ED, but no active symptoms during pregnancy. Overall, rates of postnatal depression in both of these studies are considerably higher than rates of postnatal depression in the general pregnant population, which is estimated to range from 10 to 18.8 percent (Australian Institute of Health and Welfare, 2012; O'Hara & Swain, 1996).

Due to these early role transition difficulties, women who experience disordered eating in pregnancy may struggle to in interact and bond with their newborn, placing the mother-child attachment at risk (Astrachan-Fletcher et al., 2008; Park, Senior, & Stein, 2003; Patel, Wheatcroft, Park, & Stein, 2002). This relationship may be further challenged should a woman struggle to initiate or sustain breastfeeding. While existing literature in this area is mixed, evidence from small and large population-based studies in Sweden (Larsson & Andersson-Ellström, 2003) and Norway (Torgersen et al., 2010) revealed that women with a current or lifetime history of disordered eating were significantly more likely to have ended

breastfeeding between three (Larsson & Andersson-Ellström, 2003) and six months (Torgersen et al., 2010) postpartum, compared to unaffected women. Given the complex cascade of hormonal and nutritional factors required to produce breast milk, women with EDs or subclinical disordered eating may experience difficulties sustaining breastfeeding behaviours, possibly due to a lack of milk supply (Watson et al., 2014). While the implications of limited breast milk for an infant are varied, poor infant growth has been associated with breastfeeding difficulties amongst mothers with AN (Evans & le Grange, 1995; Treasure & Russell, 1988; Waugh & Bulik, 1999).

In contrast, a study utilising the ALSPAC cohort in the UK revealed women with EDs were significantly less likely to cease breastfeeding prematurely relative to unaffected women (Micali, Simonoff, & Treasure, 2009). Watson et al. (2014) noted these differences might be due to discrepancies in diagnostic definitions, in addition to differences in sociocultural attitudes toward breastfeeding. That is, significantly fewer women reportedly breastfeed in the UK, compared to Norway and Sweden (Lande et al., 2003; Lawson & Tulloch, 1995). Other studies, however, have indicated women with EDs or subclinical disordered eating may use breastfeeding or breast milk expression as a purgatory method (Elran-Barak, Zubery, & Steiner, 2014) or prolong breastfeeding for weight control (Tiller & Treasure, 1998).

In addition to these immediate postpartum outcomes, there may be significant long-term risks for offspring of mothers with threshold EDs during pregnancy, both physically and psychologically (see Astrachan-Fletcher et al., 2008, for a full review). While literature exploring the impact of subclinical disordered eating on offspring is relatively scant, research has indicated the developing infant brain is more sensitive to a mother's diet than previously assumed (Antonow-Schlorke et al., 2011; Keenan et al., 2013; Micali et al., 2009). In the MoBa cohort, for example, women with binge eating symptoms before and during pregnancy had significantly higher overall energy intake, total fat, monounsaturated fat, and saturated fat

consumption, in addition to lower intakes of folate, potassium, and vitamin C, compared to healthy referent women (Siega-Riz et al., 2008). The researchers suggested this high fat consumption may result in fetal programming that predisposes the unborn infant to overeating and metabolic syndrome later in life (Watson et al., 2014) and internalising disorders such as anxiety (Sullivan et al., 2010). A growing body of literature has also suggested that low birth weight, which is typical of women with restrictive or purging symptom profiles, increases the risk of an infant developing cardiovascular disease and obesity in adolescence and adulthood (Abe, Minami, Ohrui, Ishimitsu, & Matsuoka, 2007; Roseboom et al., 2001; Siega-Riz et al., 2004; Thomas et al., 2012; Wren, 2011). Collectively, these findings highlight the importance of early identification and intervention to prevent both short- and long- term consequences for the mother, child, and the mother-child relationship. Antenatal care has been identified as an opportune circumstance for such symptoms to be screened for, identified, and monitored/managed appropriately (NEDC, 2015).

Perinatal Mental Health Screening and Assessment

Screening is the process of identifying individuals who display symptoms of, or may be experiencing, a particular condition or disease (Wilson & Jungner, 1968). This is achieved through application of a validated test, questionnaire, examination, or other procedure that can be performed in an inexpensive, harmless, and rapid manner (Wilson & Jungner, 1968). Screening interventions are designed to facilitate early identification and management, with the hope of reducing undesirable short- and long-term consequences (Reilly et al., 2013a). Screening tools are not, however, intended to be diagnostic. Individuals who obtain a positive screening result should receive (or be referred for) further assessment. In addition to assisting with clinical identification, screening serves an epidemiological purpose in population-based studies to elucidate the prevalence, incidence, and progression of particular conditions

(Wilson & Jungner, 1968). Screening can also be performed in a variety of ways. In some instances, a whole population group may be screened, irrespective of individual risk status. This is known as mass or universal screening. In other circumstances, only a subset of the population may be screened, particularly those identified as being in a high-risk group. This is known as selective or indicated screening (Wilson & Jungner, 1968).

Although screening tests aim to facilitate early identification and management/ treatment, not all screening tests benefit the individual being screened. Potential adverse effects of screening include incorrectly identifying healthy individuals as unwell (false positive) and identifying a number of individuals with symptoms that are unlikely to cause significant harm (overdiagnosis). As such, it is vital that any screening test, especially one that aims to identify low incidence conditions, has an excellent level of sensitivity (ability of a test to correctly identify people as unwell) and a good level of specificity (ability to correctly identify people as well; Wilson & Jungner, 1968).

The Introduction and Progression of Perinatal Screening in Australia

For the 280,000 women who give birth in a 12-month period in Australia, antenatal care is a routine component of the pregnancy experience (Australian Department of Health & Ageing, 2012). In addition to providing support and information throughout pregnancy, antenatal care seeks to monitor a woman's physical and psychological health via regular clinical check-ups, various preventative screening, and assessment of psychosocial factors. While international recommendations released by the World Health Organisation (2002) promote a package of at least four antenatal visits, women in Australia typically receive a greater level of care. On average, a schedule of 10 visits is recommended for women in their first pregnancy without complications, whereas a schedule of seven visits is suggested for subsequent uncomplicated pregnancies (Australian Department of Health & Ageing, 2012).

Assessment of a woman's risk and the need for more intensive care is, however, continually evaluated throughout the pregnancy.

Despite the prevalence of mental health conditions occurring at a similar or greater rate than most pregnancy-related complications, screening in pregnancy overwhelmingly focuses on the detection and prevention of physical illness in the mother and her unborn child (Hooper, 1996; Kapadia et al., 2015; Mayo, Melamed, Vandenberghe, & Berger, 2015; Paré et al., 2014). Undoubtedly, these physical conditions have significant and undesirable maternal and foetal effects; however, similar chronicity is also associated with untreated maternal mental health conditions (beyondblue, 2008). As noted earlier in this review, poor maternal mental health can significantly impact the emotional, social, physical, and cognitive development in infants, increase the long-term risk and incidence of chronic disease in the mother and her child, and adversely affect secure parent-infant attachment and family formation (Bauer et al., 2015; Gavin et al., 2005; Gray, 2013; Lovestone & Kumar, 1993; Meltzer-Brody & Stuebe, 2014; Oates, 2015). As such, there is a critical need for all parents/primary caregivers to have robust mental health in the perinatal period.

In 2004, Australian evidence suggested that a significant proportion of maternal mental health conditions were going undetected, undisclosed, and untreated in the perinatal period (Austin & Priest, 2004; Austin et al., 2017). In 2001, recognising the importance of mental health in the perinatal period, the Australian government funded the first phase of the beyondblue National Postnatal Depression Program (2001-2005). The comprehensive research undertaken at this time demonstrated the magnitude and impact of maternal psychosocial morbidity in Australia, particularly depression, which was found to affect up to 10 percent of women (1 in 10) during pregnancy, with this increasing to 17 percent (1 in 7) in the first-year post birth (Buist & Bilszta, 2006). Findings of this national research project also

revealed widespread acceptability of routine psychosocial screening and assessment by consumers (women/mothers) and health professionals (beyondblue, 2008).

Results of phase one, combined with seminal documents from other international organisations at the time (e.g., NICE), resulted in the development of Australia's National Action Plan for Perinatal Mental Health (2008-2010), which aimed to translate research into practice. With mental health concerns identified as one of the top three causes of indirect maternal mortality in Australia (Austin, Kildea, & Sullivan, 2007), the Australian government invested $85 million to fund the five-year National Perinatal Depression Initiative (NPDI; Highet & Purtell, 2012). The NPDI (2008-2013) represented a national approach to early identification and treatment of perinatal mental health conditions through the implementation of routine screening and access to appropriate services (Highet & Purtell, 2012). Additionally, the initiative sought to establish national guidelines for the screening and management of perinatal depression, known as the beyondblye clinical practice guidelines (2011), and engage in workforce training and development of health professionals (Highet & Purtell, 2012).

Following the success of the NPDI, universal screening of depression and anxiety, in addition to psychosocial factors that affect a woman's mental health during pregnancy, is now recommended in Australia's antenatal guidelines (Austin et al., 2011; Austin et al., 2017; Australian Department of Health & Ageing, 2012), similar to England (NICE, 2007) and Scotland (Scottish Intercollegiate Guidelines Network [SIGN], 2012). In Australia, it is recommended that clinicians administer the Edinburgh Postnatal Depression Scale (EPDS; Cox, Holden, Sagovsky, 1987) as early as practical during pregnancy and, at a minimum, repeat screening at least once later in the pregnancy (Austin et al., 2017). The EPDS is the most widely used instrument to assess symptoms of depression and anxiety in the perinatal period (Boyd, Le, & Somberg, 2005), consisting of 10-items scored on a four-point Likert

scale. The EPDS does not aim to provide a formal diagnosis; rather it intends to identify women who may benefit from further monitoring, follow-up, and support (Austin et al., 2011). According to Australia's national clinical guidelines (Austin et al., 2011; Austin et al., 2017; Australian Department of Health & Ageing, 2012), if a woman obtains a positive screen, equivalent to a score of 13 or more, monitoring and further assessment may be required, particularly if a woman has other psychosocial factors impacting her pregnancy experience. Furthermore, regardless of the EPDS score, if a woman endorses the EPDS item relating to self-harm, an immediate risk assessment must be undertaken (Austin et al., 2011; Austin et al., 2017).

While data to assess the long-term impact of the NPDI universal screening program is still being collected, a recent study using a subsample of women (N = 1,804) derived from the Australian Longitudinal Study of Women's Health revealed that women who received assessment of their mental health during pregnancy were twice as likely to receive adequate monitoring and a referral for further assessment and support (Reilly et al., 2013b). Furthermore, women who were not asked about their emotional health were significantly less likely to seek psychological assistance during pregnancy and in the postpartum period (Reilly et al., 2014). Similar findings have also been revealed in the United States (Yawn et al., 2012) and Hong Kong (Leung et al., 2011). These results highlight the crucial role of screening in not only identifying and monitoring concerning symptomatology, but also reducing stigma and assisting women to navigate the mental health care system during the perinatal period.

Challenges to Perinatal Screening in Australia

Despite the many improvements made in perinatal mental health over the past two decades in Australia, several challenges still remain (Austin et al., 2017). First, screening is not consistent and widespread; particularly in the private sector where women are 40 percent

less likely to receive screening in relation to their current emotional health. Possible explanations for this screening discrepancy in the private sector include inaccurate beliefs that affluent women are not afflicted by perinatal mental health concerns and/or time pressures in a busy clinical environment resulting in a reactive rather than proactive approach to the assessment and management of perinatal mental health (Reilly et al., 2013a). Second, there continues to be a low level of awareness and understanding of perinatal mental health in the community (Highet, Gemmil, & Milgrom, 2010) and high levels of associated stigma that has a negative impact on disclosure (Highet, Stevenson, Purtell, Coo-Calcagni, 2014). Third, there is a need for training among health professionals to increase knowledge around perinatal mental health and to improve confidence and competence in performing comprehensive and sensitive perinatal mental health screening (Highet & Purtell, 2012; Reilly et al., 2014). Fourth, referral and treatment pathways for identified cases are still somewhat limited and/or not well known by health professionals (Highet & Purtell, 2012). Lastly, at the current time mental health screening in pregnancy focuses significantly on detecting symptoms of depression and anxiety, which are considered prevalent mental health concerns during this period. Notably, the prevalence of disordered eating and threshold EDs during pregnancy is similar to antenatal depression and anxiety (Broussard, 2012; Easter et al., 2013; Lai et al., 2005; Micali et al., 2007; Pettersson et al., 2016), yet routine screening of these symptoms is not explicitly recommended or mentioned in national antenatal guidance in Australia, or internationally.

With an international focus on optimising the psychological wellbeing of women during the perinatal period and reducing mental health related morbidity (WHO, 2009), it is unclear why screening for disordered eating in pregnancy is excluded from national antenatal practitioner guidelines. Possibly, it has been incorrectly surmised that such symptoms do not affect women during pregnancy or only afflict a very small percentage of women and, as

such, do not warrant explicit screening. Alternatively, it might have been assumed that such concerns would be detected when discussing and/or monitoring healthy weight gain and eating during pregnancy. Research has revealed, however, that while clinicians working in antenatal care believe they are discussing weight and nutrition (Farrar, Butterfield, Renz, Jones, & Syson, 2013), surveys of women suggest that less than half are receiving appropriate advice and support to manage their weight and eating during pregnancy (Brown & Avery, 2012; Wiles, 1998). Reasons for this include the perceived sensitivity of such topics; frontline antenatal practitioners (e.g., doctors, midwives, and nurses) being less likely to receive comprehensive training in diet/nutrition and, therefore, having lower confidence with such topics; women having limited opportunity to discuss such topics, particularly if practitioners have decreased confidence in the area; and communication issues within the health care team (Furness et al., 2015). Traditional signs/symptoms of disordered eating can also manifest as normal symptoms of pregnancy (e.g., fatigue, emotion fluctuations, appetite changes) or they can be masked or explained by pregnancy-related ailments (e.g., pregnancy morning sickness disguising self-induced vomiting, or pregnancy-related appetite decreases concealing dietary restriction; Easter et al., 2013). Cognitive and affective symptoms can also be concealed or not easily identified without direct query or discussion. Therefore, it seems unlikely that disordered eating symptomatology would be easily detected in antenatal care unless a woman self-discloses and/or explicit physical or behavioural indicators are present.

The ability to identify or detect disordered eating in pregnancy is further complicated when attitudinal issues are considered, particularly the impact of volitional stigma. Volitional stigma is form of stigma unique to EDs and disordered eating, whereby symptoms are perceived to be “voluntary” or “self-inflicted” and, as such, under the individual’s control (Bannatyne & Stapleton, 2015; Easter, 2012). A large body of literature has revealed that due to stigma, many women experiencing disordered eating or suffering from a threshold ED are

reluctant to disclose their condition or symptoms (Franko & Spurrell, 2000; Franko & Walton, 1993; Freizinger, Franko, Dacey, Okun, & Domar, 2010; Hollifield & Hobdy, 1990; Morgan, 1997; Newton & Chizawsky, 2006; Tierney et al., 2013). Given pregnant women are expected to engage in an array of risk-aversive behaviours to ensure the health and optimal development of their unborn child, non-disclosure of disordered eating symptomatology is likely to be heightened during pregnancy, as women may be fearful that the presence of disordered eating symptomatology will elicit negative reactions from health professionals, in addition to family members and friends (Freizinger et al., 2008; Tierney et al., 2013).

While few studies have directly assessed the attitudes and opinions of antenatal providers in relation to disordered eating symptomatology during pregnancy, one study has assessed the knowledge of, and stigma toward, disordered eating and threshold EDs in a non-pregnant context. In a sample of obstetricians and gynecologists (N = 115) from four teaching hospitals in Australia and the UK, Morgan (1999) revealed that almost one third of respondents (particularly males) perceived disordered eating and EDs to be "abnormal behaviour in the context of a weak, manipulative, or inadequate personality" (p. 234), highlighting the potentially pejorative construction of EDs and disordered eating behaviours in this professional field and the likelihood of poor symptom disclosure from women due to increased fear of stigma and shame compared to other developmental periods. While exploring the impact of pregnancy for women with symptoms of AN and BN, Lemberg and Phillips (1989) revealed more than half the sample did not disclose disordered eating symptomatology to their antenatal practitioners, a concerning finding given the increased risk of undesirable fetal and maternal complications that could be managed or reduced with appropriate care and support from a specialist multidisciplinary team (Lowes et al., 2012). Likewise, in a systematic review of eight qualitative studies carried out with women

experiencing disordered eating in pregnancy, Tierney et al. (2013) revealed that, overall, women across these studies did not disclose their symptoms to antenatal clinicians.

Identifying Disordered Eating in Antenatal Care: Practitioner Beliefs and Practices

As noted earlier in this review, disordered eating in pregnancy has also been linked to numerous negative consequences such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and other obstetric and postpartum difficulties (Linna et al., 2014; Watson et al., 2014). As such, screening for disordered eating in pregnancy may facilitate early identification and management, which could mitigate associated long-term health consequences for women and children. Over the past decade, a large body of research has noted antenatal care should include regular questions regarding a woman's body weight, eating practices/attitudes, and weight control behaviour/s during pregnancy (Abraham, 2001; Bulik et al., 2007; Franko & Spurrell, 2000; Lemberg & Phillips, 1989; Micali & Treasure, 2009; Stewart et al., 1990; Wolfe, 2005). This suggestion is also supported by prominent clinical guidelines released by the National Institute for Health and Care Excellence (NICE, 2010, 2017b) and national organisations in Australia such as the National Eating Disorders Collaboration (NEDC, 2015), which highlight pregnancy as a vulnerable and high-risk period for disordered eating, similar to puberty. Despite strong support for these actions, research has revealed that embodiment of these recommendations is rare.

In Morgan (1999), 27 percent of obstetricians or gynecologists (N = 115) rarely or never inquired about EDs or disordered eating symptoms in antenatal care and only 20 percent were confident in their ability to identify a threshold ED. Around the same time, Abraham (2001) revealed that in a sample of 68 experienced obstetricians from an Australian hospital, less than half the sample inquired about disordered eating or methods of body weight and shape control, while no physician calculated pre-pregnancy body mass index (BMI). Furthermore, despite epidemiological statistics from large prospective pregnancy

cohort studies suggesting threshold EDs affect up to 7.5 percent of women during pregnancy (Bulik et al., 2007; Easter et al., 2013; Micali et al., 2007; Watson et al., 2012), with this prevalence rate potentially greater if subclinical presentations are considered, at least one third of the respondents from Abraham (2001) reported they did not believe they had treated or managed a pregnant woman with an ED in the year prior (average delivery of 125 babies a year per participant). More recently, in a study of 968 obstetricians and gynecologists in the United States, Leddy et al. (2009) revealed less than half the sample assessed ED history, body image concerns, weight-related cosmetic surgery, methods of weight control, and binging and purging behaviours. Although most physicians (90.8%) agreed EDs and disordered eating can negatively impact pregnancy outcomes, only half viewed assessment of disordered eating symptomatology as their responsibility.

Collectively, findings of these studies (e.g., Abraham, 2001; Leddy et al., 2009; Morgan, 1999) highlight the importance of encouraging antenatal providers to recognise the importance of these issues in pregnancy, and to incorporate routine screening into their practice patterns. Antenatal providers are well positioned to screen for and identify disordered eating concerns, as it is one of the rare occurrences in which women are heavily engaged in systematic and consistent healthcare, with various screening opportunities (NEDC, 2015; Ward, 2008). For example, as noted by NEDC (2015), screening opportunities include the initial pregnancy consultation (i.e., confirmation of pregnancy), various ultrasound appointments (particularly 12- and 20-weeks), the prenatal hospital admission interview, and third trimester check-ups. Each of these scenarios provides the opportunity for early detection, potentially increasing the likelihood of women receiving additional support during pregnancy, which may have protective effects for the mother and her offspring (Fornari, Dancyger, Renz, Skolnick, & Rochelson, 2014). Researchers have, however,

debated how often and under what circumstances screening and assessment of disordered eating should occur during antenatal care.

While some researchers have argued opportunistic screening should be a routine (universal) practice for all women in antenatal care regardless of presentation (Abraham, 1998; Franko & Spurrell, 2000; Lowes et al., 2012; Mitchell & Bulik, 2006) and continue throughout the course of pregnancy (Harris, 2010), others have suggested screening should be selective and only occur when indicated by certain symptoms and/or historical factors (Andersen & Ryan, 2009; Bansil et al., 2008; Hawkins & Gottlieb 2013; Ward, 2008). For instance, the NICE (2017b) guidelines pertaining to the recognition and treatment of EDs suggests screening for disordered eating should occur when an individual in any primary care context presents with: an unusually low or high BMI/body weight, rapid weight loss, disproportionate concern about body weight and shape, dieting or restrictive eating practices, unexplained gastrointestinal symptoms/pain, physical symptoms of starvation or compensatory behaviours (e.g., self-induced vomiting, use of laxatives/diuretics), psychological distress, and/or is involved in activities associated with a high risk of disordered eating (e.g., professional sport, modeling). Other researchers have detailed pregnancy-specific screening indicators, suggested clinicians should be concerned when a woman fails to gain an appropriate amount of weight across two consecutive antenatal visits (particularly during the second trimester), when a diagnosis of hyperemesis gravidarum has been given, or when there is a lifetime history of disordered eating (Bansil et al., 2008). See Table 5 for a summary of clinical presentations suggested to warrant screening.

Table 5

Summary of Presentations that Necessitate Screening (i.e., Selective Screening Indicators)

Screening Indicator
Physical/Medical
Low pre-pregnancy BMI or BMI less than 18[1, 2, 3, 4]
Lack of weight gain over two consecutive appointments (particularly during 2nd trimester)[3, 4, 5, 6, 7]
Severe weight loss or low weight in relation to stage of pregnancy[8]
Severe weight gain or excessive weight in relation to stage of pregnancy[8]
Fainting, dizziness, headaches, or excessive fatigue[8]
Unexplained gastrointestinal symptoms/problems/pain[1, 2, 3, 8]
History of oligomenorrhea or amenorrhea[3, 8]
History of infertility[3]
Diagnosis of hyperemesis gravidarum at any point [3, 5, 6] or after 20 weeks[7]
Unexplained hyperkalemia or other electrolyte imbalances[4]
Low bone density[8]
Psychological
Disproportionate concern about body weight and shape[1, 2]
Concern, distress, or preoccupation with weight gain, even when weight is within expected range[8, 9]
Negative or unusual attitudes toward food and/or eating[8]
Negative attitudes toward the unborn baby[8]
Psychological problems (e.g., diagnosis of anxiety and/or depression)[1, 2, 3, 7, 8, 9]
Depression and/or anxiety about pregnancy and/or caring for the baby[8]
History of an eating disorder[5, 6]
Low self-esteem[9]
Behavioural
Indications of food intake restriction or repeated, self-induced vomiting[1, 2, 8]
Restriction of certain foods, not advised by a clinician[8]
Avoidance of meals or changes in eating behaviour[1,8]
Evidence of substance/medication use in order to maintain body weight[8]
Self-harming or suicidal behaviour[8]
Excessive or distorted exercise patterns (or signs of distress when exercising is not possible[3, 8]
Involvement in high risk activities or interests (e.g., professional sport, dance, fashion/modeling)[1]

Note. [1]NICE (2017b), [2]Ward (2008), [3]Andersen & Ryan (2009), [4]Harris (2010), [5]Bansil et al. (2011), [6]Tierney et al. (2011), [7]Hawkins & Gottlieb (2013), [8]NEDC (2015), [9]Knoph-Berg et al. (2011).

Notably, a recent systematic review (Downe, Finlayson, Tunçalp, & Metin Gülmezoglu, 2015) cautioned that formal antenatal care has been overly focused on clinical detection and identification of actual or potential medical pathology. In response to this finding, the WHO (2016) released a series of recommendations to facilitate a positive pregnancy experience, with particular emphasis on tailored, woman-centered antenatal care, whereby appropriate preventative screening is incorporated with local knowledge and practices; relevant and timely information, and suitable emotional and psychological support. This potentially suggests that indicated/selective screening might be the preferred approach if screening for disordered eating was to be incorporated in antenatal care; however, women in previous research have indicated that routine psychosocial and mental health screening improved the therapeutic alliance with their antenatal care provider, facilitating and enhancing disclosure of concerns (Kingston et al., 2015; Reid et al., 1998). In particular, women have reported feeling most comfortable when antenatal practitioners initiate the screening process in a routine manner and when less confronting, more anonymous modes of screening are employed such as self-report modalities (Kingston et al., 2015).

Conversely, there is a widely held concern that routine mental health screening in antenatal care may cause undue psychological harm due to the risk of false-positives and stigma (Rollans, Schmied, Kemp, & Meade, 2013; Shakespeare, Blake, & Garcia, 2003). As such, it has been suggested the benefits of early detection and treatment do not outweigh the harm this may cause to the mother and/or the therapeutic relationship between a mother and her antenatal practitioner. Given these conflicting views, exploration as to whether assessment of disordered eating in antenatal care should occur, and if so, under what circumstances and using which methods, is warranted. To achieve a robust and authentic understanding of this issue, including potential advantages and limitations, exploration of practitioner and consumer perspectives is needed.

Advantages of Identifying Disordered Eating in Pregnancy

Life changes and priority shifts are thought to make pregnancy an ideal time to address and modify ingrained behaviours and thinking patterns (Wiles, 1994); however, mental health concerns often go undetected and untreated in the perinatal period (Austin et al., 2017). As noted earlier in this review, Australia's national guidelines for antenatal care (Australian Government Department of Health and Ageing, 2012) and mental health care in the perinatal period (Austin et al., 2017) have both emphasised that a woman's physical and mental health must both be a central focus in the delivery of antenatal care. Routine psychosocial and mental health screening in the perinatal period is considered vital to increase the likelihood of women receiving early intervention and management, if needed (Austin et al., 2011; 2017).

In the context of disordered eating symptomatology, the nature of this early intervention may differ depending on the severity and frequency of symptoms and level of impairment. This may include regular monitoring and early education about healthy eating to ensure a woman's caloric and nutrient intake is meeting the requirements of her own body and the unborn child (Chizawsky & Newton, 2006), preparing a woman for the numerous physical changes that pregnancy entails (Andersen & Ryan, 2009; Czech-Szczapa et al., 2015), in addition to positively reinforcing maternal weight gain and shape changes by concurrently discussing fetal growth and development (Ward, 2008). To prevent the normalisation of disordered eating symptoms, particular care should also be taken to help women differentiate between symptoms of disordered eating and changes in thoughts, feelings, and behaviours that occur as a result of a normative pregnancy experience (Chizawsky & Newton, 2006). For frontline antenatal practitioners (e.g., midwives and doctors) and potentially other allied health professionals to provide and/or facilitate such support, more frequent and longer antenatal appointments may be required for women

experiencing disordered eating in pregnancy (Harris, 2010; Lowes et al., 2012; NICE, 2004; Ward, 2008). In cases where there is risk of harm to the mother and/or unborn child, specialist multidisciplinary treatment incorporating medical monitoring, high-risk obstetric management, structured nutritional intervention, and psychotherapy may be necessary (Harris, 2010; Lowes et al., 2012; NEDC, 2015). As such, strong collaboration and communication between health care providers is vital and referral to a specialist ED service may be warranted to facilitate appropriate clinical intervention (Bulik et al., 2007; Lowes et al., 2012; NEDC, 2015).

Although there is clear guidance for monitoring and managing the physical risks associated with disordered eating in pregnancy, there is a paucity of literature regarding appropriate psychological intervention during this specific developmental stage (Crow et al., 2008; Soares et al., 2009; Tierney et al., 2013). For screening and early detection to be of greatest benefit, it is important that effective evidence-based interventions exist (Public Health England, 2015; Wilson & Jungner, 1968). At the current time, international guidance relating to the clinical management of antenatal mental health concerns (NICE, 2017a) has indicated that women with disordered eating symptomatology during pregnancy should be offered psychological intervention, with recommendation that such interventions are based upon best-practice guidelines NICE (2017b) for the recognition and treatment of EDs (see Table 6)

Table 6

Overview of the NICE (2017b) Psychological Treatment Recommendations for EDs

ED Subtype	Recommended Psychological Treatment/s

AN	• First line treatment:
	  ○ Eating disorder focused cognitive behaviour therapy (CBT-ED) – individual treatment
	  ○ Maudsley anorexia nervosa treatment for adults (MANTRA) – individual treatment (with family/friend involvement)
	  ○ Specialist supportive clinical management (SSCM) – individual treatment
	• If all 3 first line treatments are ineffective or contraindicated:
	  ○ Eating disorder focused focal psychodynamic therapy (FPT) – individual treatment

BN	• First line treatment
	  ○ Bulimia nervosa focused CBT guided self-help program (with brief face-to-face intervention focused on adherence)
	• If guided self-help is unsuitable, contraindicated, or ineffective after 4 weeks:
	  ○ Individual CBT-ED

BED	• First line treatment
	  ○ Binge eating focused CBT guided self-help program (with brief face-to-face intervention focused on adherence)
	• If guided self-help is unsuitable, contraindicated, or ineffective after 4 weeks:
	  ○ Group based CBT-ED
	• If group CBT-ED is not available, is unsuitable, or the person declines:
	  ○ Individual CBT-ED

OSFED	• Use the first line treatment for the ED it most closely resembles

The recommendation to treat disordered eating in pregnancy as per the NICE (2017b) ED guidelines does, however, present several issues. First, the NICE (2017b) guidelines make no direct reference to the treatment of subclinical disordered eating concerns, other than the suggestion that treatment of OSFED should be guided by the ED it most closely resembles. Second, although the interventions outlined in the NICE (2017b) ED guidelines

have demonstrated efficacy in non-pregnant populations, it is unknown whether this efficacy translates to pregnancy (Crow et al., 2008; Soares et al., 2009; Tierney et al., 2013). It is reasonable to suggest that treatment approaches may require modification in the perinatal period. Despite these limitations, it is widely believed that psychological intervention and support for disordered eating in pregnancy would be beneficial (Harris, 2010; Lowes et al., 2012; NEDC, 2015; Ward, 2008) and a continued and compassionate relationship with an antenatal practitioner would result in a positive difference (Easter, 2015). Given the positive influence pregnancy reportedly imparts on a woman's motivation for change, future research investigating the treatment of disordered eating symptomatology in pregnancy is warranted. Early intervention during pregnancy could prevent progression into the postpartum period where symptoms are often exacerbated (Crow et al., 2008), in addition to mitigating or reducing undesirable foetal and maternal consequences that could have a detrimental and enduring effect (Fornari et al., 2014; NICE, 2017a).

Summary and Gaps in Literature

This review of literature has highlighted that while pregnancy may permanently or temporarily improve disordered eating symptoms, the powerful biopsychosocial event may also represent a period of increased vulnerability, potentially exacerbating or triggering the onset of disordered eating behaviours and/or cognitions. Although current estimates suggest that up to one in four women (28%) may experience symptoms of disordered eating during pregnancy, such symptoms are often undetected and undisclosed due to fear of stigma, poor knowledge and awareness of disordered eating symptomatology during pregnancy, difficulty distinguishing disordered eating from normative pregnancy symptoms, and potentially unsuitable assessment instruments.

Antenatal practitioners are, however, well positioned to identify symptoms of disordered eating in a gentle and non-confrontational manner, and support women to engage

in appropriate treatment, if required. Pregnancy is thought to represent a unique window of opportunity to make significant and sustainable symptom changes. It is also posited that antenatal treatment may prevent or mitigate the risk of relapse or symptom exacerbation postpartum (Astrachan-Fletcher et al., 2008; Edelstein & King, 1992). Development of a standardised, pregnancy-specific screening instrument, sensitive to the eating and weight-related changes that occur during pregnancy, may facilitate symptom detection and early intervention. For this to occur, however, the conceptualisation of disordered eating in pregnancy and the distinction from normative pregnancy symptomatology must first be established.

The Current Thesis

Given the possible frequency with which disordered eating is occurring in pregnancy and the suggestion that undesirable maternal and fetal consequences associated with threshold EDs may extend to subclinical variants (Crow et al., 2008; Harris, 2010), the overarching aim of this PhD was to improve the identification of disordered eating in pregnancy. To achieve this aim, a series of research questions were developed:

1. How does disordered eating manifest in pregnancy? (Chapter 3)
    1.1. Is this similar or distinct from disordered eating in a non-pregnant context?
    1.2. Does this perception differ between experienced health professionals and women with a lived experience?
2. How is disordered eating symptomatology distinguished from pregnancy-appropriate symptomatology? (Chapter 3)
    2.1. Where is the threshold between the two constructs and how do experienced health professionals determine this distinction?
    2.2. Does this perception differ between experienced health professionals and women with a lived experience?

3. Should screening for disordered eating occur in antenatal care and, if so, should this occur on a universal or selective basis? (Chapter 3)
4. What instruments are currently available to screen for disordered eating in pregnancy? (Chapter 4)
    4.1. If available, are these tools psychometrically sound and validated for use in pregnancy? (Chapter 4)
    4.2. If not, could a standardised and psychometrically sound pregnancy-specific screening instrument for detecting disordered eating be developed? (Chapters 4 and 5)

Outline of Studies

To answer these research questions, a mixed methodological approach was employed with four sequential studies conducted between 2015 and 2017. A brief description of each study is outlined below, including the research questions (RQ) addressed in each study.

To clarify the conceptualisation of disordered eating in pregnancy, Studies 1 and 2 used a Delphi methodology (explained in Chapter 2) to explore the views of professionals and consumers (i.e., women with a lived experience) pertaining to the symptom expression of disordered eating in pregnancy (RQ 1) and how this is distinguished from pregnancy-appropriate symptomatology (RQ 2). Studies 1 and 2 also aimed to elicit and understand views on the assessment of disordered eating in antenatal care (RQ 3) to determine whether assessment should occur in antenatal care, and if so, under what circumstances and using which methods. Study 3 aimed to systematically identify and review objective measures of disordered eating to determine their suitability for use in pregnancy populations (RQ 4). Study 4 aimed to develop a pregnancy-specific disordered eating screening instrument based on the findings of Studies 1 and 2, in addition to exploring the psychometric properties (reliability and validity) of the instrument using a sample of pregnant women in Australia.

CHAPTER TWO

The Delphi Technique: Methodological Review

Chapter Overview

The aim of this chapter is to provide contextual background in relation to the Delphi technique. The Delphi technique is the primary methodology used in Studies 1 and 2 (Chapter 3) to understand professional and consumer perspectives on the expression and assessment of disordered eating in pregnancy. The first half of the chapter provides a brief overview of the Delphi technique and the underlying theoretical basis, while highlighting the importance of consensus approaches in evidence-based practice. The second half of this chapter provides a practical guide to the Delphi process and its variants, highlighting the importance of research-practice translation. The chapter concludes with a description of the approach adopted by the two Delphi studies in the current thesis (Studies 1 and 2).

Overview of the Delphi Technique

The Delphi technique (Dalkey & Helmer, 1963) is a formal methodology which has traditionally aimed to achieve consensus among a group of experts when an accepted knowledge base is absent or lacking (Graham et al., 2003; Hardy et al., 2004; Mead & Moseley, 2001; Sumison, 1998). While the formal technique originated from a series of studies conducted by the RAND Corporation in the 1950s, which aimed to forecast and explore the impact of technology on warfare (Dalkey & Helmer, 1963), certain researchers report the conceptual implementation can be traced back to the oracle of Delphi in ancient Greek history (Baker, Lovell, & Harris, 2006). It is reported the oracle used a number of informants to pronounce the ‘truth’, enhanced by information or data from various sources (Kennedy, 2004). Similarly, although the formal Delphi technique was first used in the 1950s, informal consensus methods had reportedly been used in health care for many years prior (Murphy et al., 1998).

In a broad sense, the technique involves several iterative questionnaires (rounds) to canvass and organise the opinions of a group of individual experts (panelists), who typically remain anonymous to avoid power imbalances and the phenomenon of group think (Williams & Haverkamp, 2010). The panel moderator provides structured feedback in between each round, usually summaries of the quantitative results and qualitative themes from the previous rounds. This multi-stage procedure generally continues until a certain level of consensus is reached (Hasson et al., 2000) or, in more recent years, the 'stop' criterion is met (i.e., total number of rounds set *a priori*; Holey, Feeley, Dixon, & Whittaker, 2007). Panel members typically engage in an ongoing reflection process in response to the intermittent feedback, which then often (but not always) results in panel members converging toward consensus on a topic. While the key elements of the Delphi process are described in greater depth later in the chapter, a broad overview of the methodology can be seen in Figure 1 below.

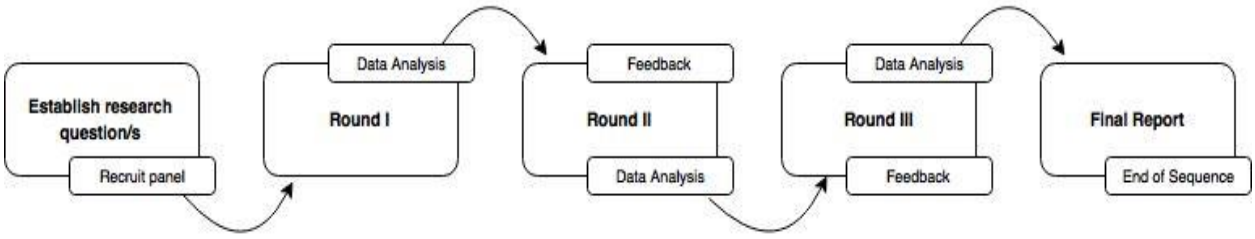

Figure 1. Brief overview of the Delphi technique process.

The logic behind the consensus process in the Delphi technique is somewhat informed by the wisdom-of-crowds effect (Surowiecki, 2004), which posits that under most circumstances a collated series of imperfect estimates by a group will produce a better outcome than a single estimate by an individual expert; an effect that has been found to be robust under various conditions and circumstances (Davis-Strober et al., 2014; Jönsson et al., 2015). In understanding the origin of these beliefs, the wisdom-of-crowds effect has been traced back to a competition estimating the weight of a butchered ox at an English country fair in the early 20th century, during which Francis Galton analysed the distribution of the 787

estimates in the competition. Galton's analyses revealed that the median estimate was remarkably accurate (within 0.8%), indicating that aggregating the guess of a group produced a reliable estimate of the outcome. In this scenario, the crowd could be considered 'wise' (Surowiecki, 2004); however, there are times when crowds are not wise and group pressures can lead to irrationals decisions, as seen in the phenomenon of group think (Jorm, 2015).

Surowiecki (2004) has proposed there are four main conditions that must be met for groups to be wise:

1. Diversity of expertise – better quality decisions will be produced when a group is heterogeneous, as opposed to homogeneous;
2. Independence – to ensure group members are not influenced by each other, individuals must be able to make their decisions independently and anonymously;
3. Decentralisation – group members must be autonomous and work in a decentralised manner;
4. Aggregation – a mechanism to coordinate and aggregate the group's expertise must be in place.

When considered and appraised in reference to the wisdom-of-crowds literature, the Delphi method appears to incorporate many of the conditions that Surowiecki (2004) proposes lead to groups being wise. For example, in the Delphi method, the group responds anonymously to a series of questionnaires over time (independence of decisions), panelists operate autonomously (decentralisation), and decisions are shared to the group by the panel facilitator following aggregation and statistical summarisation of the panel's responses (aggregation). Although the remaining condition for a group to be wise (diversity of expertise) is not necessarily a requirement of the Delphi technique, researchers must consider diversity when selecting panel members to ensure optimal decision making is promoted.

According to Page (2007), in complex decision making, individuals use inherent predictive models to produce an estimate. For example, as noted by Jorm (2015), an individual's predictive model in estimating the weight of the butchered ox might be: "This ox appears to be about five times my size – I weigh 80 kilograms, therefore the ox must weight approximately 400 kilograms" (p. 888). As such, for a crowd to produce optimal predictions the individuals in the crowd must have good predictive models and there must be diversity between these models (Page, 2007). A crowd where only a limited number of predictive models are available will perform much worse than a crowd with a diverse range of models. Even more problematic, if all members of the crowd use the same predictive model to generate the same estimate or judgment, the crowd will be no better at predicting an outcome than a single individual (Page, 2007). As such, although diversity is not a compulsory component of the Delphi technique, it is an element that should be strongly considered when selecting the panel (Jorm, 2015).

Why Consensus Methods are Needed

Over the past two decades, there has been a strong move toward evidence-based medicine, a practice whereby clinical and policy decisions are consciously and explicitly informed by the highest quality evidence at the time (Greenhalgh, Howick, & Maskrey, 2014; Sackett & Rosenberg, 1995). The quality of information has typically been assessed and defined by the Levels of Evidence statements. According to these statements, the strongest form of evidence is a systematic review of randomised controlled trials, with lower quality evidence arranged underneath to form a hierarchy. Notably, in several of these evidence categorisation schemes, expert opinion is invariably low on the hierarchy and is often listed as one of the weakest forms of scientific evidence, as seen in the Joanna Briggs Institute Levels of Evidence for Effectiveness (Joanna Briggs Institute & University of Adelaide, 2013).

Researchers have argued that expert opinion should not automatically be considered an inferior method; particularly given expert opinion has been the primary methodological approach in establishing many of the existing evidence-based medicine tools (Jorm, 2015). For example, not only were the Levels of Evidence statements developed based on expert consensus, other tools such as the Cochrane Handbooks for Systematic Reviews of Interventions (Higgins & Green, 2011), the Consolidated Standards of Reporting Trials (CONSORT) statement (Begg et al., 1996), and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009), also used expert consensus methods in their development. As noted by Jorm (2015), when expert consensus methodologies are used in this manner, they become a foundation for the development of other methodologies.

When expert consensus is considered in the broader context, it also acts as a fundamental component of science. For example, expert consensus is often used to determine which methodologies are considered appropriate, where funding should be allocated, which manuscripts should be published, and who should be admitted to prestigious societies of experts (Jorm, 2015). More specifically, citation metrics may also reflect consensus of discipline regarding the importance of a publication. According to Jorm (2015), science can be conceptualised as what the community of experts in a particular field consider to be the truth at a given time; however, this consensus is flexible and is likely to change as new information emerges.

The rate at which consensus develops and changes is, however, somewhat dependent on the field of investigation. For example, in physical sciences, a single piece of strong evidence may be sufficient to change expert views. Conversely, in other sciences dealing with multiple systems, many of which are highly complex (as in mental health), the

consensus process is much slower and formal mechanisms to access consensus are often needed, such as the Delphi technique (Jorm, 2015).

Furthermore, although it would be ideal for all decisions and recommendations in health care to be made on the basis of evidence derived from rigorously conducted empirical studies, in some areas, sufficient research evidence may not exist and it is reasonable to assert that sufficient evidence is unlikely to ever exist (Rowe & Wright, 1999). As such, the ability to make effective decisions in circumstances where there is insufficient or contradictory information has resulted in increased use of formal and informal consensus methods (Hasson, Keeney, & McKenna, 2000). The intention of these consensus methods is not to challenge or replace statistical or model-based procedures, but to act as an option or solution in situations where statistical methods are not practical or possible (Rowe & Wright, 1999). In this respect, consensus methods make best use of available information, whether this is scientific data or the collective wisdom of the group members, potentially creating new understanding of this information.

Uses of the Delphi Methodology in Mental Health Research

Although the Delphi technique was developed for technological forecasting in the 1950s (Dalkey & Helmer, 1963), the methodology has been used in multiple settings as a way of gaining consensus and/or clarity on topics, issues, or definitions within a particular field. The technique has an extended history of application within psychology (e.g., Norcross, Hedges, & Prochaska, 2002; Norcross, Koocher, & Garofalo, 2006; Spinelli, 1983; Thielsen & Leahy, 2001) and mental health research. In particular, the Delphi technique has been utilised to clarify diagnostic issues (Dawson, Rhodes, & Touyz, 2015; McFarlane, Owens, & del Pozo Cruz, 2016; Noetel, Dawson, Hay, & Touyz, 2017) and clinical performance indicators (Mittnacht & Bulik, 2015) within the field of EDs.

With application in mental health research, various uses for the Delphi technique have emerged. To ascertain the types of consensus decisions, Jorm (2015) systematically explored 176 Delphi articles between 2000 and 2015. Overall, four main consensus uses were revealed:

1. Making estimations where evidence is incomplete (e.g., what is the global prevalence of dementia?)
2. Making predictions (e.g., what types of interactions with a person experiencing suicidal ideation will reduce their risk of suicide?)
3. Determining collective values (e.g., what should the performance indicators for mental health care be?)
4. Defining foundational concepts (e.g., how should we define relapse in schizophrenia?)

It should be noted that although these broad categories provide a simple overview of the range of uses a Delphi study can elicit, consensus decisions in many studies often integrate more than one category. For example, a study may not only seek to define a foundational concept, but also to make predictions in a particular area.

Variations of the Delphi Technique

Several variations of the Delphi technique currently exist and are accepted based on the design, purpose, and anticipated response fatigue of the study, in addition to the response stabilisation and the number of diminishing returns (Dalkey et al., 1972; Hasson, Keeney, & McKenna, 2000). These variations include increasing or decreasing the number of iterative rounds, incorporating pre-existing content from literature into questionnaires, and utilising heterogenous or homogenous samples. Some of these differing forms are known as the ‘modified Delphi’ (McKenna, 1994) in which fewer iterative rounds are employed often by pre-populating the first round questionnaire, the ‘policy Delphi’ (Crisp et al., 1997) whereby

consensus and dissensus is explored in relation to a preferred future, and the 'real-time Delphi' (Beretta, 1996) in which single iterated rounds are not required and panelists continue to re-complete questionnaires as many times as necessary in a specified period via an online server, with instant feedback provided throughout.

With consideration of these variations, the traditional conceptualisation of the Delphi technique as a method to obtain reliable consensus from a group of experts using a series of iterative questionnaires interspersed with controlled feedback (Dalkey & Helmer, 1963) has been broadened to encompass the various interpretations of the technique. For example, Linstone and Turoff (1975) have described the Delphi technique as a method of group communication that is structured and thereby effective in allowing group members, as a whole, to consider and attempt to resolve complex issues. Similarly, Reid (1988) has conceptualised the Delphi technique as a method for systematically collecting and aggregating informed judgment from a group of individuals on specific issues or questions. Overall, regardless of which definition is accepted, the general aim of the Delphi technique is to establish, predict, and explore the attitudes, needs, and priorities of a group (Hasson & Keeney, 2011).

The Delphi Process

While numerous variants of the Delphi technique exist, Jorm (2015) outlines several key steps, which are detailed below and depicted graphically in Figure 2 at the end of this section. Notably, carrying out each of the Delphi steps requires decision making by the research team based on the aim/s of the research.

Framing the research question/s. As with any empirical research, the first step of the Delphi process is to have a clear and coherent research question that can be answered by the methodology. One distinguishing quality of Delphi studies is that often the research

questions posed cannot be answered by other methodologies, or there would be considerable difficulty in doing so.

Selecting an expert panel. Unlike most empirical research, the Delphi technique uses non-probability, purposive sampling to recruit a group of individuals to form an "expert" panel; however, the term expert is contentious and ambiguous within literature. In a traditional sense, an expert is defined as an individual with strong knowledge about, and experience with, a particular issue or topic (Keeney et al., 2001). In this respect, the individual may be considered a "specialist" or "informed" (McKenna, 1994). This broad definition does present issues, namely how knowledge and experience are quantified or measured. It has been suggested that professional qualifications/registrations could qualify an individual's knowledge level; however, this does not guarantee expertise. Similarly, although duration in one's professional practice appears to be proxy for expertise, it does not ensure this (Baker, Lovell, & Harris, 2006).

To avoid any misinterpretations that can occur with the term expert, some Delphi studies have opted to avoid the term altogether (Lai, Flower, Moore, & Lewith, 2015). Other researchers recommend complete transparency in relation to the criteria used to include participants in the panel (Trevelyan & Robinson, 2015). As individual research teams must develop these inclusion criteria, there is considerable variability between studies. Some indicators of expertise utilised in previous Delphi studies have included (but are not limited to): years of professional practice or research in an area (Bovopoulos et al., 2016; Dawson et al., 2015; Mittnacht & Bulik, 2015; Noetel et al., 2017; Williams & Haverkamp, 2010; Yap et a., 2014), appointment of distinguished academic positions (Noetel et al., 2017), publication of peer-reviewed journal articles and/or books in the field under investigation (Addington et al., 2013; Dawson et al., 2015; Noetel et al., 2017; Williams & Haverkamp, 2010), presentation at peer-reviewed national conferences (Williams & Haverkamp, 2010),

membership in special interest groups (Dawson et al., 2015), and/or distinguished professional acknowledgement via invitations to closed professional societies or fellowship endowment (Dawson et al., 2015; Noetel et al., 2017). Most studies only require one criterion to be met to allow inclusion; however, some require at least two (Trevelyan & Robinson, 2015).

While the inclusion indicators outlined above relate to professional expertise, increasingly other types of expertise are being recognised as valuable to include in Delphi studies, namely consumer and caregiver advocates. In these situations, the inclusion criteria for participants may be less clear and often requires an expression of interest, followed by screening to confirm lived experience or advocacy roles. Snowball sampling may also be needed to increase recruitment (Jorm, 2015). Panelists recruited via this method must be advised to avoid discussions about the study during the Delphi process to preserve the integrity of the methodology. Overall, the type of expertise required in a Delphi panel will depend on the question under consideration (Jorm, 2015). In some circumstances, consumer consensus may be the primary outcome sought (e.g., when determining service users' perspectives of particular treatments). Alternatively, there are some topics that only a professional panel can adequately provide opinion on, for example, how to segment the hippocampus (Boccardi et al., 2015).

To ensure diverse opinions are generated, sometimes the perspective of both consumers and professionals may be desired (e.g., how to apply mental health first aid in the workplace; Bovopoulos et al., 2016). This is consistent with the wisdom-of-crowds literature, which suggests that groups often make better decisions when there are a variety of predictive models available (Bantel, 1993; Page, 2007). Existing literature supports the joint inclusion of professionals and consumers/caregivers in a panel, in addition to recommending diversity of professional expertise (i.e., including professionals from various professions; Jorm, 2015).

In other studies, professional and consumers/carers are in independent parallel panels and results are then triangulated for diversity of opinions or, less frequently, consensus is required in both panels. Notably, studies that have utilised independent parallel panels have revealed a high level of agreement between professionals and consumers when examining endorsement correlation rates (ranging from .71 to 92), despite varying types of expertise (Cairns et al., 2015; Reavley et al., 2012, 2013; Ross et al., 2014a; Ross et al., 2014b). It is important to note, however, that diversity in a panel does not guarantee expertise; panel members must still be assessed to ensure they possess relevant expertise (Jorm, 2015). Furthermore, at times, a homogenous panel may be more appropriate, particularly when an area is highly specialised (Trevelyan & Robinson, 2015).

In addition to deciding on the type of panel to utilise and developing clear inclusion criteria, it is recommended that research teams establish a clear sampling strategy to identify potential panelists (Jorm, 2015). While this strategy will differ between studies, examples in previous research have included systematic reviews/searches of relevant literature (Addington et al., 2013; Yap et al., 2014), websites or databases of professional societies (Yap et al., 2014), broad online searches (Mittnacht & Bulik, 2015), contributors on consumer websites (Ross et al., 2014a); approaching advocacy groups/organisations (Bond et al., 2017; Bovopoulous et al., 2016); direct contact (Bovopoulous et al., 2016; Williams & Haverkamp, 2010) and snowball sampling techniques (Ross et al., 2014a).

Determining the size of the panel. The sample size for Delphi studies varies considerably depending on the research topic and specific situation (Atkins, Tolson, & Cole, 2005). In general, there is a paucity of firm guidance directing researchers toward an appropriate panel size (Jorm, 2015), with literature suggesting anywhere from 10 to 50 panel members is appropriate (Linstone & Turoff, 2002). While the robustness of a Delphi study is largely dependent on the strength or quality of the expert panel, rather than size of the panel,

greater stability is achieved with larger panel sizes as the responses of individual panel members have less of an impact on the consensus outcome (Jorm, 2015; Okoli & Pawloski, 2004). For example, if an 80 percent consensus threshold is established for a panel of only 10 individuals, each individual will represent 10 percent of the overall consensus threshold, a percentage that can lead to considerable fluctuations to endorsed items (Jorm, 2015).

Attempting to determine panel size stability, one previous study in health care quality and safety utilised bootstrap sampling, revealing that a panel of 23 experts produced stable results (Atkins et al., 2005). While this finding represented a move toward quantifying the minimum sample size required, it is unclear whether the panel stability demonstrated in this study can be generalised to other fields and areas. Overall, in the area of mental health research, it is recommended that Delphi panels aim to have at least 20 members to achieve response stability (Jorm, 2015). Researchers must also account for the likelihood of attrition across rounds by recruiting a greater number of participants than required to ensure the final sample is equal to or greater than 20.

Determining the number of iterations/rounds. Although the classic Delphi method reportedly utilised a four round process (Hasson et al., 2000), research in recent years has favoured a three round process, often decided a priori (Hasson et al., 2000). In some instances, only two rounds are used; however, this does present issues for determining stability across rounds (Murphy et al., 1998). When deciding on the number of rounds to implement, various factors need to be considered including the amount of time available to the researchers, the planned structure of the first round questionnaire (i.e., open-ended questions vs. pre-populated items), panel fatigue, and attrition. As panel attrition is likely to increase with each additional round, research has indicated use of three rounds is optimal in minimising panel fatigue, but also ensuring meaningful data is obtained (Hasson et al., 2000).

Developing the Round I questionnaire. In the classical Delphi approach, the first round questionnaire consists of open-ended questions, generating qualitative data to develop the second round questionnaire. In other situations, researchers may collect first round data in a 'real world' setting (e.g., interviews and focus groups) using inductive analysis (Trevelyan & Robinson, 2015). However, in recent years modified approaches have been increasingly used, with researchers conducting systematic literature searches and using thematic or content analysis to generate key ideas and then pre-populate items in the first round questionnaire (Crowe et al., 2015).

This modified approach has several benefits to consider, namely increased time efficiency and improved utilisation of each round to assess stability. Although critics argue this modified process may miss key opinions, solutions, or arguments that would be generated via open-ended questions (Hasson et al., 2000), integrating open-ended questions and comment boxes throughout the first round questionnaire can mitigate this (Green, Jones, Hughes, & Williams, 1999; Trevelyan & Robinson, 2015). Previous research has suggested the modified Delphi technique is perceived by panel members to be cooperative and equally as effective as the original technique when executed in a flexible (i.e., allowing panel suggestions/feedback), yet rigorous manner (Eubank et al., 2016; Graefe & Armstrong, 2011; Gustafson, Shukla, Delbecq, & Walster, 1973).

As the size of the first round questionnaire varies considerably depending on the topic and research question/s under consideration, Delphi studies with longer first round questionnaires often benefit from organising or grouping items into main themes, making it easier for panel members to complete the questionnaire and to identify omissions (Jorm, 2015). Once the items have been developed, the researchers must then decide on the most appropriate response format. In most instances this takes form as a Likert scale to allow the extent of agreement within the panel to be determined (Jorm, 2015); however, there may be

situations where other response formats are required (i.e., single selection multiple choice, item rankings, dichotomy or category selections).

Notably, there is limited guidance in terms of the optimal number of response categories, with studies using Likert scales ranging from four (Smith, Grant, & Lyttleton, 2012) to 11 (Fernandes et al., 2013) points. Previous research has revealed that Likert scales with four points or less are generally associated with poor reliability and discriminating power (Preston & Colman, 2000). While this may lead researchers to conclude that Likert scales with a greater number of response points (Lozano, García-Cueto, & Muñz, 2008), an excess number of response points can produce inconsistencies in category interpretation and, as such, misleading results (Jones & Loe, 2013). It has been suggested the optimal number of Likert scale response categories lies somewhere between four and seven (Lozano et al., 2008), with most Delphi studies in the field of mental health utilising five-point Likert scales (e.g., Bond et al., 2017; Bovopoulos et al., 2016; Dawson et al., 2015; Hart et al., 2014; Mittnacht & Bulik, 2015; Noetel et al., 2017).

Deciding how much information to give panelists to aid their judgments. When administering the Delphi questionnaires across the rounds, particularly the first round questionnaire, some researchers elect to provide the panel with additional information (e.g., brief reviews of available literature, definitions of key concepts) to consider when making their judgments and ratings (Jorm, 2015). Conversely, other researchers decide not to provide additional information and leave it to each panel to draw upon their own sources of expertise (Jorm, 2015).

Whether additional information should be included is, again, at the discretion of the researchers. In making this decision, research teams must consider whether evidence is available to summarise, and the quality of such evidence. Additionally, the research team needs to remain cognisant of the type of judgment the panel is being asked to make and

whether providing evidence will bias the panel. For example, providing a review of evidence when the panel is asked to make a value based judgment may not be appropriate (Jorm, 2015).

Administering the Round I questionnaire. As the Delphi process requires panel members to make independent judgments, there is no need for panel members to meet face-to-face. As such, questionnaires can be administered via postal mail or online using a secure web-server. In traditional Delphi studies, paper-based communications were common (Green et al., 1999); however, it has been reported this approach often places significant administrative burden on the panel moderator when gathering, organising, and compiling participant responses (Cole, Donohoe, & Stellefson, 2013). For Delphi panel members, the effort required to return questionnaires via postal mail may also be deterrent in agreeing to participate or continuing with the study. These issues are likely to be amplified in the case of international Delphi participants, as paper-based questionnaires may take considerable time to reach the participant and to be returned. As other Delphi participants cannot progress without the responses of the entire panel, long wait times in between questionnaire rounds (particularly when international colleagues are involved) may result in frustration and diminished interest in the study (Donohoe & Needham, 2009), resulting in greater attrition across rounds (Sinha, Smyth, & Williamson, 2011). As such, in the case of paper-based surveys, researchers must choose whether to limit the parameters of the Delphi panel to local areas, thereby affecting diversity, or risk attrition by seeking greater panel diversity in wider geographic areas.

Conversely, online questionnaire administration enables experts from all geographic locations to be included more efficiently, potentially increasing the diversity of expertise within the panel (Jorm, 2015), reduces the costs associated with postal mail, and allows participation to be monitored via online survey platforms (Cole et al., 2013). However, online

administration does require participants to have access to a stable Internet connection and possess computer literacy skills. As more than 80 percent of households in developed nations had a stable Internet connection at the end of 2015 and, on average, 81 percent of adults in these developed nations used the Internet (Organisation for Economic Cooperation and Development [OECD], 2017), issues regarding connectivity may be less problematic in developed countries. However, Internet availability does not necessarily guarantee computer literacy. In exploring this, an OECD skills study with 215,942 individuals across 33 countries (close to 5,000 from each country) with Internet connectivity revealed that 60 percent of individuals had at least basic computer literacy skills, defined as competency in basic technology applications such as an Internet browser or email software, and the ability to acquire information from these programs using explicit criteria and search functions (OECD 2017). Overall, to ensure successful completion of the Delphi process, a research team must consider the circumstances and skills of the planned Delphi panel to select an appropriate administration method (Hasson et al., 2000).

Analysing the Round I questionnaire. Unlike cross-sectional survey methodologies, which analyse responses following the completion of data collection, the Delphi technique requires ongoing analysis to achieve consensus. In the classical Delphi method, which utilises open-ended questions to elicit the viewpoints of the panel in Round I, analysis of the Round I questionnaire would be qualitative using thematic or content technique to identify key themes and then construct appropriate quantitative items for Round II. In the modified Delphi approach, which uses existing literature to pre-populate items combined with open-ended questions to elicit panel feedback, analysis of the Round I questionnaire consists of both quantitative and qualitative methods. These quantitative methods usually involve basic frequency analyses and some measure of central tendency (e.g., mean, median, or mode) and dispersion (e.g., standard deviation or interquartile range [IQR]; Murphy et al., 1998). In

general, mode and median tend to be preferred over mean as it is more robust to the effect of outliers (Murphy et al., 1998) and appropriate for ordinal data (Manikandan, 2011). As standard deviation does not necessarily apply to ordinal data, some researchers have argued that providing the standard deviation is misleading and the IQR should be reported instead (Hasson et al., 2000; Murphy et al., 1998).

***Consensus*.** In determining panel agreement for each item, a clear definition of consensus is required; however, there is no single definition of consensus. Again, it is the responsibility of the research team to establish a clear consensus criterion. Various approaches have been utilised, including: formal measures of agreement (e.g., Kappa statistic, Cronbach's alpha, interclass correlation coefficient), the degree of uncertainty around a point estimate, stability of group responses, and the proportion (mostly percentage) of participants agreeing with a particular viewpoint (Black et al., 1999; Graham, Regehr, & Wright, 2003; Hsu & Sandford, 2007; Lindstone & Turoff, 2002). In a recent systematic review of 100 English language Delphi studies, the most common approach to defining consensus was use of a percentage agreement (Diamond et al., 2014). This finding is also consistent with an earlier systematic review of 80 Delphi studies, which revealed the most frequently used approaches were percentage agreements and median scores above a pre-determined threshold (Boulkedid, Abdoul, Loustau, Sibony, & Alberti, 2011).

When using percentage agreement approaches to consensus, researchers must also establish and clearly specify a pre-determined agreement threshold. How the consensus agreement threshold is defined differs greatly between studies and may depend on several factors, including: the nature of the question being asked, the characteristics of the panel (i.e., professional vs. consumer), number of panels involved (i.e., single vs. multiple), the number of items (small pool vs. large pool), and the potential implications of the study (Jorm, 2015; Keeney, Hasson, & McKenna, 2006). In a systematic review of 100 Delphi studies, Diamond

et al. (2014) revealed the median threshold for determining consensus was 75 percent (range: 50 to 97 percent). This threshold has also been used in previous Delphi studies in the area of EDs (e.g., Dawson et al., 2015).

Administering subsequent questionnaires and providing feedback to the panel. Across the Delphi rounds, panel members are given a series of questionnaires to complete. At the beginning of each new round, panel members are given feedback in relation to the collective responses of the group, both quantitative findings and qualitative themes. For each item, feedback on the frequency distribution of the panel's responses, the appropriate measure of central tendency and dispersion, and the percentage of endorsement is given. Visual feedback in the form of bar charts or graphs is also thought to assist the panel with interpretation (Ward, Stebbings, Sharman, Cherkin & Baxter, 2014). In some studies, panel members are reminded of their individual responses. The purpose of this is to allow panel members to compare their individual responses to those of the panel; however, not all Delphi studies employ this individual feedback mechanism due to the suggestion that it can potentially undermine the 'wisdom-of-crowds' effect (Iqbal & Pipon-Young, 2009; Lorenz et al., 2011).

Another important consideration in the Delphi process is identifying which items will be included in subsequent rounds. Some researchers elect to remove items that meet the pre-determined consensus threshold or rejection/removal threshold, even if this occurs during the first round (as in a modified Delphi when items are pre-populated). In this scenario, only items that approach consensus are included in the following round, in addition to any new items based on feedback from the first round (Bond et al., 2017; Bovopoulos et al., 2016; Jorm, 2015; Noetel et al., 2017; Yap et al., 2014). While this cross-sectional approach may reduce panel fatigue and potentially minimise attrition (Keeney, Hasson, & McKenna, 2011), it does not allow the stability of the panel's responses to be assessed (Greatorex & Dexter,

2000; Holey et al., 2007). Ideally, all items should be rated at least twice to ensure stability of responses. After this, items that meet the consensus or rejection thresholds can be removed, and only those that approached consensus are re-rated in subsequent rounds (Trevelyan & Robinson, 2015).

Reporting the results. Due to the large volume of data that can be produced because of multiple questionnaire rounds, communicating the results of a Delphi can be somewhat complex. According to Jorm (2015), one of the most straightforward approaches is to generate a list of accepted items, particularly when the number of accepted items is relatively small. In circumstances involving many accepted items, the most feasible approach may be to arrange the items under relevant thematic headings and qualitatively summarise the findings to assist with interpretation. In both situations, it may also be helpful to utilise a flowchart to demonstrate the course of items across the multi-stage process (Jorm, 2005).

Importance of Implementing Findings

As with most empirical research, simply reporting the findings of a Delphi study is unlikely to produce any direct practical benefit. However, as Delphi studies generally deal with research questions related to clinical/practice needs, it has been argued that translation of research into practice is potentially easier (Jorm, 2015). As such, it is important that researchers intending to carry out a Delphi study develop an implementation strategy during the planning phase. Jorm (2015) also highlights that if a research team has no intention or capacity to implement the findings, then the researchers should seek to consult and partner with an individual, team, and/or organisation that can carry out this important function.

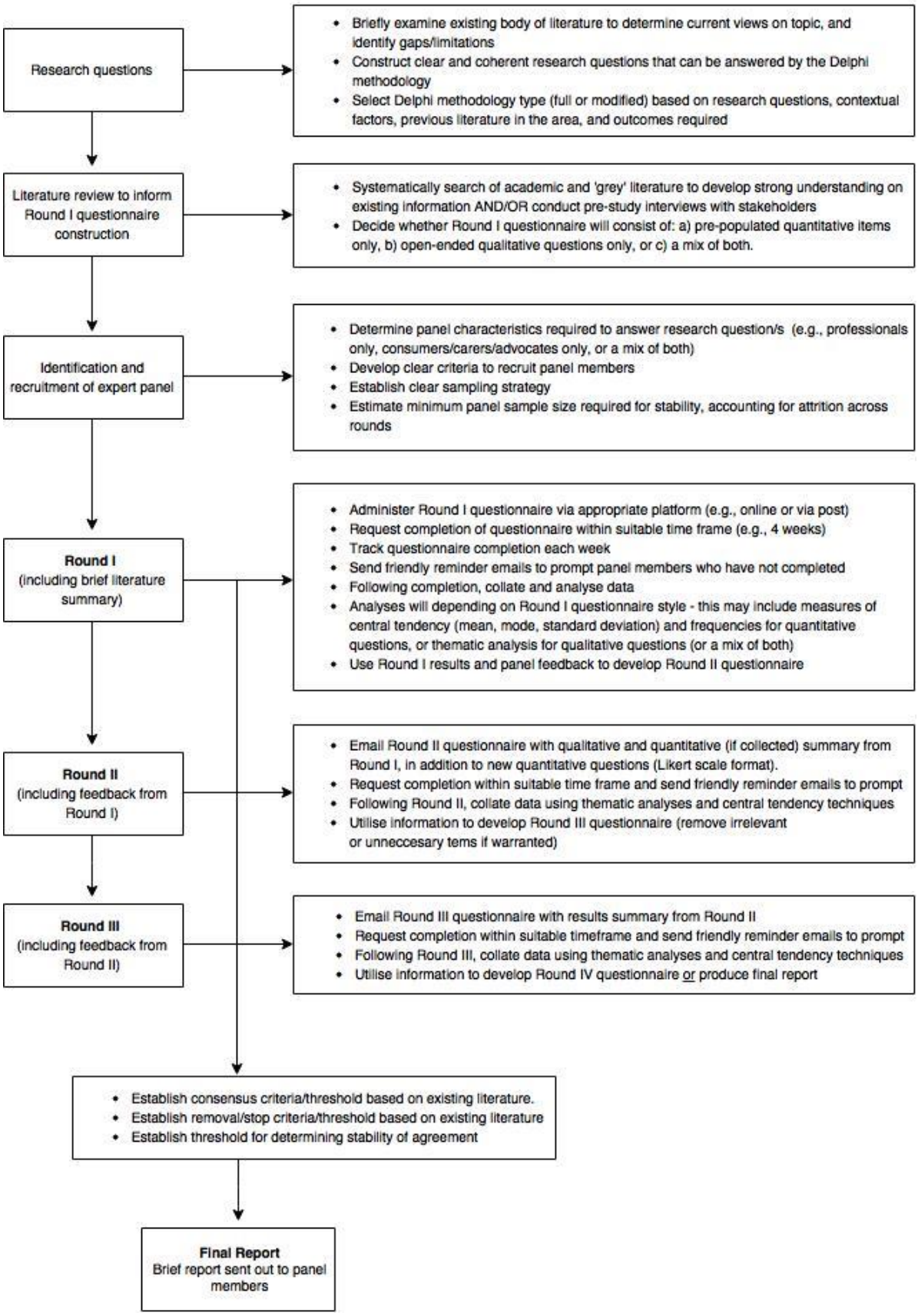

Figure 2. Detailed overview of the Delphi process (adapted from Hughes et al., 2015)

Application of the Delphi Methodology to the Current Project

Based on the literature reviewed, the Delphi technique was considered an appropriate methodology to assist in achieving the overarching aim of the current project. That is, to understand and improve the identification of disordered eating in pregnancy. The Delphi methodology was perceived to be a crucial component in the potential instrument pre-development phase given the symptomatology of disordered eating in pregnancy has not been explicitly discussed in previous literature and it cannot be confidently assumed that symptoms in a non-pregnant context would generalise to pregnancy, especially due to the clinical overlap between disordered eating and the experience of pregnancy.

While a traditional focus group methodology could have been employed to generate perspectives on this topic, this would not have enabled a consensus process to be achieved, the scope of the focus group would have been limited to local clinicians and researchers, and the data may have been influenced by power imbalances and group think (Jorm, 2015; Williams & Haverkamp, 2010). As such, the Delphi technique was considered the most appropriate methodology due to its mixed design. Moreover, some researchers have argued the incorporation of qualitative and quantitative data in the Delphi technique allows researchers to obtain a robust understanding of a phenomenon under investigation (Iqbal & Pipon-Young, 2009), while also facilitating the consensus process.

The primary intention of using the Delphi technique was not only to define a foundational concept such as the definition of disordered eating in pregnancy, but also to determine collective values in relation to the distinction between disordered eating and pregnancy and whether antenatal screening for disordered eating should occur and, if so, using which methods. To answer these research questions, a modified Delphi approach was utilised with two independent panels, one made up of ED professionals and the other comprised of consumers with a lived experience of disordered eating in pregnancy. The

purpose of using a dual panel approach was to include various forms of expertise to ensure diverse opinions could be generated and all perspectives could be considered if instrument development was necessary later in the project. Although previous studies have combined professionals and consumers into a single panel (Ross et al., 2014b), the current project separated professionals and consumers into two independent panels to allow questions/items to be tailored to the specific panel. As such, two modified Delphi studies were conducted. Despite this, there was strong consistency in the main questions asked across the two panels, enabling triangulation of results to occur. A graphical overview of this triangulation process is depicted in Figure 3. Additional study details and processes are described in Chapter 3.

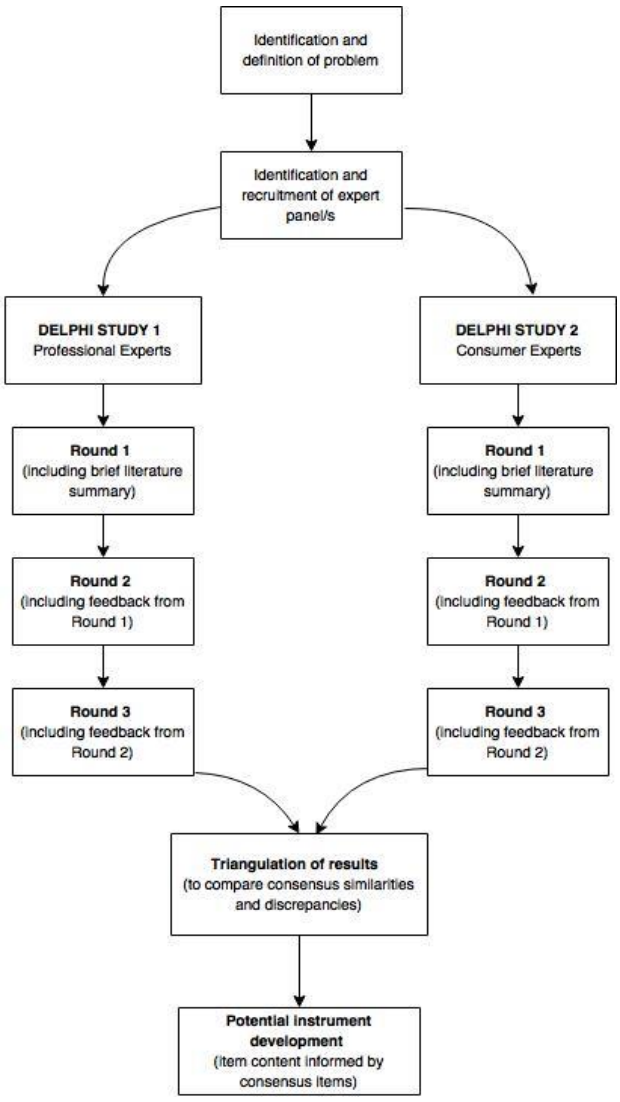

Figure 3. Graphical overview of the Delphi process implemented in the current project

Strengths of the Delphi Methodology

One of the main advantages of the Delphi technique is that it is often able to answer research questions that may not be possible or feasible with alternate methodologies, particularly quantitative types (Delbecq et al., 1975). This includes scenarios where epidemiological or experimental data are absent, incomplete, or not applicable to the problem of interest, or when consensus of values is needed (Jorm, 2015). The simple, but flexible technique enables group communication with individuals in geographically diverse locations (Fish & Busby, 1996), an advantage over other group-based methodologies such as focus groups. It is also cost effective, as panelists are not required to meet in person (Okoli & Pawlowski, 2003), with this lack of face-to-face interaction overcoming communication barriers such as reluctance to express unpopular views, to disagree with one's colleagues, or alter one's opinions (Barnes et al., 1987).

Limitations of the Delphi Methodology

Similar to most methodologies, the Delphi technique is not without limitations. Many of these limitations are described throughout the chapter; however, a few additional limitations are outlined below. First, it has been suggested the consensus reached may not reflect true consensus, but rather a compromise position or forced consensus (Hasson et al., 2000; Mitroff & Turoff, 1975). This issue can be reduced to some extent by researchers measuring consensus and/or stability using a combination of statistical methods; being explicit about (and not confusing or mixing) the terminology and criteria for consensus, agreement, and stability/reliability; in addition to panel moderators practicing reflexivity when providing controlled feedback (i.e., considering how his/her own reactions, feelings, motives may influence the interpretation of results, particularly qualitative feedback; Trevelyan & Robinson, 2015). Second, although it is proposed the methodology minimises group think, it has been argued that the structured feedback process can result in a group

think mentality, whereby panelists who hold discrepant views from the collective view their responses as "incorrect" as there is no opportunity for panelists to elaborate on their views (Goodman, 1987; Hasson et al., 2000; Walker & Selfe, 1996). This anonymity may also produce less ownership of ideas.

Third, despite being a consensus methodology, there is a seeming lack of agreement or consensus on certain methodological components (e.g., the definition of consensus); however, this has improved in recent years with publication of systematic reviews and strong recommendations. Fourth, due to the number of variants of the Delphi methodology, determining the reliability and validity of the methodology is reportedly challenging (McKenna & Keeney, 2008); however, recent reviews have recommended methodological transparency to mitigate this (Trevelyan & Robinson, 2015). Lastly, it is important to note that the existence of consensus does not necessarily mean the correct answer, opinion, or solution has been revealed (Hasson et al., 2000), as it cannot be guaranteed that results would be identically replicated in a different panel of experts. As such, the results of a Delphi do not necessarily reflect the agreed upon 'truth' in a particular area (Ludwig, 1997).

Conclusion

Overall, the Delphi technique is a simple, yet unique methodology that has been widely applied in mental health research when an accepted knowledge base is absent or lacking. When used in a robust and rigorous manner, the Delphi technique can contribute to advances in the field of interest or produce new interpretations of existing evidence (Jorm, 2015). Considered in the context of the current thesis, use of the Delphi technique may assist in clarifying the expression of disordered eating in pregnancy and the distinction between normative pregnancy-appropriate symptomatology, in addition to exploring whether agreement could be reached on the assessment of disordered eating in pregnancy.

CHAPTER THREE

Clarifying the Definition and Assessment of Disordered Eating in Pregnancy: Understanding Professional and Consumer Perspectives

Chapter Overview

This chapter details the methodology and results of the two modified Delphi studies conducted in the first phase of the overall project. These studies investigated whether consensus could be reached on the expression and assessment of disordered eating in pregnancy (Research Questions 1, 2, and 3). In the first Delphi study, the panel was comprised of individuals with professional expertise. The panel in the second Delphi study was comprised of consumer advocates. Although previous studies have combined professionals and consumer advocates into a single panel, the groups were separated into independent panels in the current project to allow questions and items to be tailored to each panel, if required. As there was a high level of consistency in the main questions answered by the two panels, comparison of findings was possible. As such, this chapter concurrently describes the methodology and results of both Delphi panels.

Rationale and Objectives

Pregnancy is a significant biopsychosocial event that often marks the beginning of a new stage in a women's lifespan development (Bulik et al., 2007). The transition to motherhood entails a multitude of rapid changes to a woman's body, eating patterns, social functioning, and self-identity, most of which are largely outside her control (Darvill et al., 2010; Larsson & Andersson-Ellström, 2003). Adjusting to these morphological, endocrinological, and psychological changes may be a relatively uncomplicated process for some women; however, for other women, the adjustment may be more challenging (Easter, 2015). These changes, combined with the age-related vulnerability of a woman's prime childbearing years to eating disturbances, means pregnancy may represent a period of

increased risk for the onset, resurgence, or exacerbation of disordered eating symptomatology, even for women with no history of such symptoms (Andersen & Ryan, 2009; Astrachan-Fletcher et al., 2008; Bulik et al., 2007; Easter et al., 2013; Harris, 2010; Hawkins & Gottlieb, 2013; Knoph Berg et al., 2011; National Eating Disorders Collaboration [NEDC], 2015; Soares et al., 2009; Tiller & Treasure, 1998; Ward, 2008). While the symptomatology of disordered eating is well understood in a non-pregnant context, the presentation and manifestation of symptoms in pregnancy is less clear. Clarifying the precise features of disordered eating in pregnancy is warranted given the increased physical and psychological risks to the mother and unborn child across the perinatal period (Linna et al., 2013; Watson et al., 2014).

As noted in Chapter 1, disordered eating is often described as a range of unhealthy eating behaviours and cognitions that can negatively impact an individual's emotional, social, and physical wellbeing (American Psychiatric Association, 2013). Typically, the distinction between disordered eating and a threshold ED has been the degree of severity and frequency of symptomatology, with disordered eating occurring at a lesser frequency and lower level of severity (NEDC, 2017). The experience of pregnancy is often accompanied by a range of abnormal eating-related behaviours and attitudes such as food cravings, increases or decreases in appetite, dietary changes, inconsistent eating patterns, food aversions, and nausea and vomiting. Despite these behaviours being normal within the context of pregnancy and potentially explained by hormonal fluctuations, changes in sensory perception, and maternal and/or fetal nutritional needs (see Orloff & Hormes, 2014), many of these pregnancy-appropriate changes overlap with, and possibly mask, disordered eating symptomatology. For example, binge eating behaviours could be confused with "eating for two" (when a pregnant woman consumes a greater quantity of food with the misperception that large caloric increases are necessary to nourish herself and her unborn child, or uses such

reasoning to explain consumption of large quantities of food or certain food types such as highly processed products), persistent pregnancy sickness could be explained by purging, and changes in dietary preferences and/or reduced appetite could be equated to dietary restriction. Antenatal practitioners may, therefore, struggle to identify disordered eating in pregnant women, particularly when symptoms fluctuate between alleviation and exacerbation depending on the course and stage of pregnancy (Easter, 2015). In many instances practitioners also lack the required training and confidence for such identification (Leddy et al., 2009).

Given the possible frequency with which disordered eating is occurring in pregnancy, combined with the indicated negative maternal and fetal consequences associated with threshold and subclinical variants of EDs (Crow et al., 2008; Harris, 2010), development of a psychometric instrument that reliably identifies disordered eating symptoms during pregnancy is warranted (Easter et al., 2013; Pettersson et al., 2016). For such an instrument to be developed, however, clarifying the precise features of disordered eating in pregnancy and how they are distinguished from pregnancy-appropriate symptomatology is required. To date, this has not been systematically explored. Furthermore, researchers have debated under what circumstances screening and assessment of disordered eating should occur during antenatal care. As such, there were three main objectives of the study:

1. To determine whether consensus could be reached on the symptom manifestation of disordered eating in pregnancy;
2. To determine whether consensus could be reached on the factors that are important in distinguishing disordered eating from pregnancy-related symptomatology;
3. To determine whether consensus could be reached in relation to the antenatal assessment of disordered eating in pregnancy (i.e., should it occur and, if so, under what circumstances and using which methods).

Method

To achieve these objectives, a modified Delphi approach (Dalkey & Helmer, 1963; McKenna, 1994) was used to establish expert consensus on key issues. As outlined in Chapter 2, the modified Delphi was selected over the classical Delphi due to the increased time efficiency of three rounds and the improved utilisation of each round to assess stability. Research has indicated use of three rounds is optimal in minimising panel fatigue and attrition, but also ensuring meaningful data is obtained (Hasson et al., 2000). The current study followed the protocol outlined by Jorm (2015), which has been replicated widely in the field of mental health, mostly for the use of guideline development. Ethics approval via the Bond University Human Research Ethics Committee (#15278) was obtained prior to the commencement of the study (see Appendix A).

Participants (panelists)

To ensure diverse opinions could be generated and all perspectives considered when answering the research questions (Powell, 2003), two independent panels were recruited. International clinicians and researchers with expertise in the field of disordered eating, particularly in relation to pregnancy and/or women's health (hereafter "professionals") formed one panel. Women who identified with a lived experience of disordered eating in pregnancy (hereafter "consumers" or "consumer advocates") formed the second panel.

Professional panel. Professional expertise in the current study was defined as meeting one of the following criteria: 1) established interest and expertise in the treatment of disordered eating, preferably within the context of the perinatal period, and/or women's health; 2) distinguished contribution to the field of EDs/disordered eating as evidenced by i) award of Fellowship status by the Academy for Eating Disorders (AED), ii) appointment as Associate Professor or Professor in the field of EDs/disordered eating and/or women's health, iii) more than 10 years experience working in the field of EDs/disordered eating and/or

women's health, or iv) publication of peer-reviewed journal article(s) and/or book(s) focused on EDs/disordered eating and/or women's health in the perinatal period.

As interdisciplinary treatment is a well-established and preferred practice in the field of EDs (APA, 2013; Garner & Garfinkel, 1997), a heterogeneous sample of professionals was desired. Panel members were recruited from English speaking developed countries. Researchers were identified through authorship of relevant articles during a systematic review of literature, while clinicians were identified via online searches, membership of special interest groups, and professional network suggestions. Additionally, AED Fellows with relevant clinical or research interests, as listed on the AED website, were contacted.

Potential panelists for the professional panel were contacted via email with an invitation to participate in the study. This email outlined the rationale and purpose of the study, how the results would be used, and the procedure of a Delphi (i.e., number of rounds, qualitative and quantitative components, anonymity of the process). A flowchart of recruitment and participation can be found in Figure 4. Of the 80 emails that were delivered, there was a 44 percent response rate. This response rate is similar to recently published Delphi studies in the area of EDs and disordered eating (e.g., MacFarlane et al., 2016; Mittnacht & Bulik, 2015; Noetel et al., 2017).

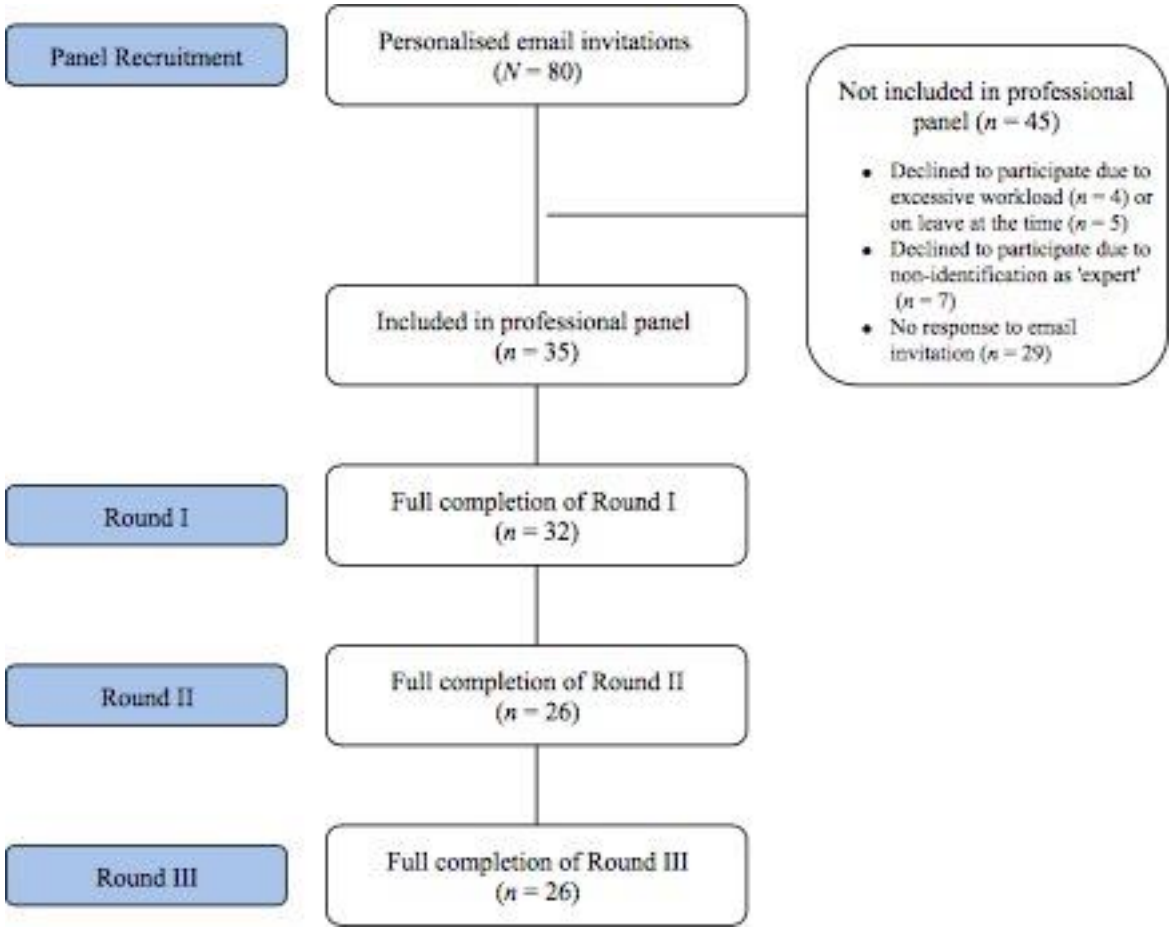

Figure 4. Recruitment and participation overview for the professional panel.

Consumer panel. Unlike recruitment for the professional panel, it was not possible for the researchers to employ purposive invitation-based sampling for the consumer panel due to ethical reasons (i.e., clinicians or advocacy groups respecting confidentiality). As such, expression of interest recruitment was utilised, similar to the recruitment of consumer advocates in other Delphi studies (e.g., Ross et al., 2014). This was achieved by posting advertisements on online pregnancy and parenting forums, in addition to targeted advertising on social media platforms. Women who identified with an experience of eating disordered thoughts, feelings, or behaviours in pregnancy (of any severity), and were interested in participating in the study, were asked to contact the primary researcher and briefly detail their experience. This primarily occurred via email; however, telephone contact details were also provided.

As one of the main aims of the Delphi process was to clarify the symptomatology of disordered eating in pregnancy, the inclusion criteria for the consumer panel were broad. As such, any woman who described eating-, body-, or exercise- related behaviours, attitudes, or thoughts that were distressing or caused functional impairment during pregnancy was invited to participate in the study. Women who disclosed a medical condition that may have produced such symptoms (e.g., hyperemesis gravidarum) were not invited to participate. Women were invited to participate regardless of symptoms being active or inactive at the time of recruitment. Exercise-related behaviours, thoughts, or feelings were included as such symptoms can be used as a compensatory mechanism to control weight/shape/size in the presence or absence of problematic eating behaviours, in addition to these behaviours causing affective distress (APA, 2015).

As part of the invitation process, a detailed explanatory statement was emailed to the participant outlining the rationale and purpose of the study, how the results would be used, and the procedure of a Delphi (i.e., number of rounds, received anonymity of the process, the function of the panel, and the nature of feedback to be received). Overall, the present study aimed to recruit a panel of at least 20 consumer advocates, consistent with Delphi panel size recommendations outlined by Jorm (2015). A flowchart of recruitment and participation can be found in Figure 5. Of the 22 consumers who were invited to participate, there was an 86.4 percent participation rate in Round I.

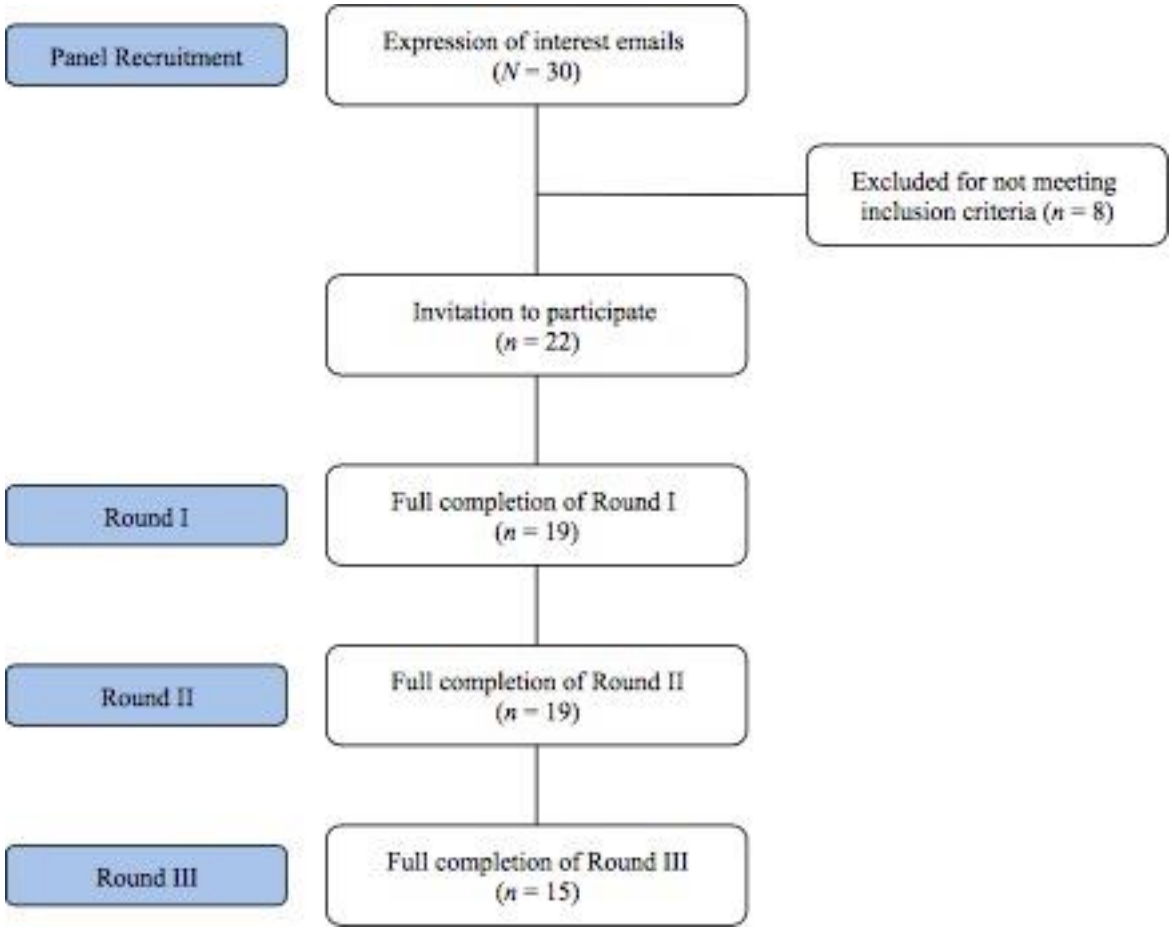

Figure 5. Recruitment and participation overview for the consumer panel.

Due to the time commitment required over a six-month period, the consumer advocates were compensated a total of $100.00 in Coles electronic gift vouchers. As the likelihood of attrition increases incrementally over the course of a Delphi study (Hsu & Sandford, 2007), these vouchers were distributed in a manner designed to enhance panel retention. That is, a $25.00 gift voucher was emailed following the completion of the Round I questionnaire, another $25.00 gift voucher was emailed following the completion of the Round II questionnaire, while a $50.00 gift voucher was emailed following the completion of the Round III questionnaire. Similar compensation values for consumer have been employed in previous Delphi studies (Reavley, Ross, Killackey, & Jorm, 2013)

Procedure

Data were collected across three questionnaire rounds (6 across the two panels) between March and November 2016 using a secure, online survey platform (Qualtrics). Panelists were given four to five weeks to complete each questionnaire round, with reminder emails sent twice during each questionnaire completion period. The questionnaire was formatted to allow each participant to save his/her responses and return at a later day/time. It should be noted there was no direct or indirect contact between the two panels.

Consensus. As noted in Chapter 3, unlike traditional methodologies that analyse the data following the completion of a study, the Delphi technique requires ongoing analysis to facilitate the consensus process. In accord with Diamond et al. (2014), consensus was defined as at least 75 percent agreement (i.e., ratings of *important* and *very important,* or *agree* and *strongly agree*) on an individual item. All items were rated at least twice (i.e., in Rounds I and II) prior to the decision to include ($\geq$ 75% agreement), re-rate in Round III (50-74% agreement) or remove (< 50% agreement). Items suggested at the end of Round I were automatically rated in Rounds II and III to obtain two rounds of data. Items were evaluated independently in each panel, and then compared at the end of the study. At the beginning of each new round a summary of the group results from the previous rounds was provided. Items that reached the 75 percent consensus agreement threshold were highlighted for panelists using bolding and asterisks.

Round I. Consistent with a modified Delphi approach, a comprehensive literature search of both academic and grey literature was conducted between October and December 2015 to inform the content of the initial questionnaire. Grey literature refers to both published and unpublished material that is not available commercially and is generally not subject to peer review (Lawrence, Houghton, Thomas, & Weldon, 2014; Schöpel & Farace, 2010). This can include informal communications (e.g., blogs), newsletters, pamphlets, information

sheets, and reports, to name a few (Lawrence et al., 2014). Key terms were used to locate relevant websites, journal articles, reports, clinical guidelines, books (including diagnostic criteria), booklets, and training manuals. Consistent with Bond et al. (2017), the grey literature search was conducted using Google Australia, Google UK, Google USA, and Google Books, while the academic literature search was performed using PubMed and PsycINFO databases. The key search terms used were: (eating disorders OR disordered eating) in pregnancy; (manage* OR support* OR treat*) (disordered eating OR eating disorders) in pregnancy; (defining OR symptoms of) disordered eating in pregnancy; (screening OR assessment OR identification) of (disordered eating OR eating disorders) in (pregnancy OR antenatal OR perinatal OR maternity care).

Similar to previous Delphi studies, the first 50 items in each search were retrieved and reviewed for relevance after the removal of duplicate sources (Bond et al., 2017; Langlands et al., 2008a). Previous research has revealed the quality of resources rapidly declines after the first 50 search hits (Kelly, Jorm, Kitchener, & Langlands, 2008). As recommended by Bond et al. (2016), to minimise the influence of searching algorithms on Google, several steps were undertaken: 1) the history in Google's search settings was routinely cleared to minimise the influence of previous searches, 2) care was taken to ensure the primary researcher was not logged into any Google-related accounts (e.g., Gmail) that may utilise demographic details to target searches or information; 3) location features that may bias information presented were disabled and the 'any country' function on Google's searches was de-selected to ensure only local pages in each search region were shown.

Sources were included if they related to EDs/disordered eating specifically in the context of pregnancy and addressed the key areas under consideration. Pertinent information from each source was categorised thematically according to the areas of investigation in a spreadsheet by the primary researcher. When a search hit generated a website landing page

with multiple hyperlinks, all links were reviewed. The information in the spreadsheet was then utilised to pre-populate the items in the Round I questionnaire. Overall, 200 sources were used to develop the Round I questionnaire (see Table 7).

Table 7

Summary of Sources that Contributed to the Development of the Round I Questionnaire

Source type	Number included	Example/s
Websites (general educational materials, pamphlets, news articles, forums)	72	https://www.thewomens.org.au/health-information/pregnancy-and-birth/mental-health-pregnancy/eating-disorders-in-after-pregnancy/ http://www.cci.health.wa.gov.au/docs/ACF383.pdf
Empirical journal articles	84	Easter et al. (2013), Tierney et al. (2013)
Clinical guidelines or reports	18	National Eating Disorders Collaboration (2015)
Conference proceedings	3	Burton (2014)
Theses	6	Taborelli (2015)
Books	17	American Psychiatric Association (2013) Hendrick (2006)

In consultation with the research team (one with expertise in the Delphi methodology and public health, one with expertise in EDs, one with expertise in dietetics and nutrition, and one with expertise in psychometric instrument development), the items generated from these 200 sources were reviewed and reworded (if necessary) for consistency and to avoid repetition, whilst retaining the original meaning. The primary researcher met with each member of the research team on several occasions to finalise the Round I questionnaire (see Appendix B), which resulted in three main sections. In section one, panelists were asked to indicate the extent to which they agreed that an item reflected a sign or symptom of disordered eating in pregnancy on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). A total of 61 symptoms were presented to both panels for rating in Round I. In section two, panelists were asked to indicate how important certain factors were in

distinguishing disordered eating symptomatology from pregnancy-appropriate symptomatology (foci items) on a 5-point Likert scale (1 = *not important* to 5 = *very important*). A total of 32 foci items were presented to both panels for rating in Round I. In section three, panelists were asked to indicate whether screening should be a routine component of antenatal care (i.e., occur for every woman), only occur when indicated by presenting signs/symptoms and/or historical factors, or not occur at all. Panelists were asked to rate each option on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). Panelists were also asked to review and rate the suitability of 12 potential assessment methods for identifying disordered eating in antenatal care. These methods were again rated on a 5-point Likert scale (1 = *not suitable at all* to 5 = *very suitable*). In addition to suitability, the consumers were asked to rate their level of comfort with each method on a 5-point Likert scale (1 = *not comfortable at all* to 5 = *very comfortable*).

While the items in these sections were consistent across both panels, additional explanations were provided in instances of jargon. Each section included a short, balanced summary of existing literature, appropriate to the knowledge level and experience of the respective panels. Care was taken to ensure that any literature provided was balanced and did not present as a means of persuasion. This was achieved through a process of questionnaire cross-checking by each researcher. Panelists were, however, encouraged to draw upon their own experiences when responding to each item.

To allow rich data to emerge for subsequent questionnaire round development, open-ended questions were included in the Round I questionnaire to facilitate and elicit feedback and suggestions for additional items in each section. Round II and III also included open-ended text boxes; however, use of these was limited to panelists contextualising responses (if required) or providing feedback to the panel moderator if there was difficulty answering a question. Prior to administration, the final version of the Round I questionnaire was piloted

on 10 colleagues unconnected to the study (5 academic researchers and 5 clinicians) to ensure adequate face and content validity.

Following the completion of the Round I questionnaire, the responses from both panels were pooled and analysed separately in SPSS Version 23 using measures of central tendency (mean and mode), dispersion (standard deviation), and frequency. Qualitative feedback elicited from the open-ended comments boxes in both panels were downloaded and transferred into separate Word processing documents and analysed using thematic analysis. Common themes in each panel were identified and cross-coded by two independent researchers to ensure accuracy. These comments were then translated into new quantitative items to be included in Round II of the respective panel, provided the ideas had not been included in the Round I questionnaire and were relevant to the scope of the project.

It should be noted that although the professional and consumer panel were recruited concurrently, there was a delay in receiving the Round I responses of four consumer panel members due to technology difficulties. To prevent significant attrition from the professional panel, a decision was made to send out the Round II questionnaire for the professional panel, while waiting for the consumer responses to be returned. The outcome of this decision was that Round I item suggestions from the professional panel (8 new symptom items, 1 new foci item) could be incorporated into the Round II questionnaire of both the professional and consumer panel; however, the Round I item suggestions from the consumer panel (20 new symptom items, 1 new foci item) could only be incorporated into the Round II questionnaire of the consumer panel (i.e., the professional panel did not rate new items suggested by the consumer panel). This also meant that items ratings were evaluated independently in each panel. In other words, the two panels operated independently of each other until the end of the study when items that reached consensus in both panels were compared.

Round II. To obtain at least two rounds of data for each item, all items from the Round I questionnaire were re-rated in Round II, regardless of whether consensus had been reached in the independent panel or not. For the professional panel, 9 additional items (8 new symptom items, 1 new foci item) were introduced in Round II. An additional 21 items (20 new symptom items, 1 new foci item) were introduced in Round II for the consumer panel. For both panels, administration of the Round II questionnaire was identical in terms of instruction and format to the Round I questionnaire; however, the Round II questionnaire included a summary of the respective panel results from Round I at the beginning of each section. This summary included both central tendency scores for each item and a summary of the qualitative feedback for the respective panel. Items that reached the 75 percent consensus agreement threshold were highlighted for panelists using bolding and asterisks. Panel members were again invited to provide qualitative feedback during Round II via the text boxes.

Following the completion of Round II, a similar data collation and analyses process was performed on the data. Items that met the pre-determined consensus agreement threshold ($\geq$ 75%) were considered endorsed and, as such, were not included for re-rating in Round III. The remaining items were either re-rated in Round III (50-74% agreement in Round II) or removed (< 50% agreement in Round II). As most items in each section met either the consensus or removal criteria by the end of Round II, only items that were new to Round II required re-rating in Round III. No new symptom or foci items were introduced in Round III.

Round III. As the Round III questionnaire was much briefer than previous rounds, follow-up questions based on qualitative panel feedback were incorporated in Round III. These follow-up questions differed slightly for the professional and consumer panels. Both panels were asked to determine the broad frequency at which symptoms might be considered "disordered" in pregnancy. These symptoms were framed as "a significant influence of body

weight and shape on self-evaluation in the presence of any compensatory behaviour aiming to prevent/reduce pregnancy-related weight gain AND/OR the presence of binge eating episodes/behaviours that occur and are followed by feelings of guilt or shame". Response options included *once per month, once per fortnight, once per week,* and *twice per week.* Panelists were asked to select one response. The purpose of this question was to identify a broad proxy that may assist clinicians to distinguish disordered eating from normative pregnancy experiences.

The professional panel was also asked to rate the suitability of an existing screening tool, in addition to indicating the ideal item length of an instrument to be used in clinical practice. The consumer panel was asked to provide further feedback on some of the assessment methods perceived to be both practical and comfortable. Following the completion of Round III, responses were again pooled and analysed using measures of central tendency (mean and mode), dispersion (standard deviation), and frequency, and then compared to the consensus agreement threshold.

Results

Panel Demographics

Professional panel. A total of 32 experts were recruited, with 26 completing all three rounds (81.3%). Overall, the final sample consisted of 23 women and three men from geographically diverse areas, with an average of 19.08 years (*SD* = 11.56) respective professional experience and 14.42 years (*SD* = 10.97) specialisation in the field of EDs/disordered eating. Seven panel members also identified as AED Fellows, a status that recognises distinguished contributions in the area of EDs. See Table 8 for additional panel details.

Table 8

Demographic Characteristics of the Professional Panel (N = 26)

Demographic variable	Frequency (%)
Residing country	
Australia	12 (46.2%)
United States	6 (23.1%)
United Kingdom	4 (15.4%)
Canada	2 (7.7%)
Sweden	2 (7.7%)
Highest level of education	
Doctorate / PhD	19 (73.1%)
Masters Degree	4 (15.4%)
Postgraduate Degree (unspecified)	2 (7.7%)
Undergraduate Degree	1 (3.8%)
Professional field	
Psychology / Psychiatry	21 (80.1%)
Dietetics	4 (15.4%)
Obstetrics	2 (7.7%)
Midwifery	1 (3.8%)
Professional activities	
Researcher also involved in clinical practice	11 (42.3%)
Clinician with no research activities	8 (30.8%)
Researcher with no current clinical practice	4 (15.4%)
Clinician with some research involvement	2 (7.7%)
Other (e.g., semi-retired – activism/editor)	1 (3.8%)

Consumer panel. A total of 19 women were recruited, with 15 completing all three rounds (79.0% retention rate). The age of the final sample ranged from 23 to 43 years (M = 45.62 years, SD = 12.08). At the time of recruitment, five women were pregnant (31.2%), one had given birth within the past six months (6.3%), one had given birth within the past year (6.3%), seven had given birth within the past two years (43.8%), and one had given birth

within the past three to five years (6.3%). In exploring the index pregnancy that disordered eating was experienced in, ten women (66.7%) reported an experience of disordered eating in only one pregnancy, with 70 percent noting this was experienced in their first pregnancy (*n* = 7). Five women (33.3%) reported experiences of disordered eating in multiple pregnancies, including their first pregnancy. For most of the panel, disordered eating was experienced during a planned pregnancy (80.0%). Of the five women who were pregnant during the study, all had given birth previously and all reported experiencing disordered eating in their previous and current pregnancy. Additional demographic information is outlined in Table 9.

Table 9

Demographic Characteristics of the Consumer Panel (N = 15)

Demographic variable	Frequency (%)
Ethnicity	
Caucasian/European	13 (86.6%)
Aboriginal/Torres Strait Islander	0 (0.0%)
Asian	1 (6.7%)
Pacific Islander	1 (6.7%)
Hispanic/South American	0 (0.0%)
Middle Eastern	0 (0.0%)
African or North African	0 (0.0%)
Highest Level of Education	
Did not finish high school	0 (0.0%)
Year 12	0 (0.0%)
TAFE/Private College	6 (40.0%)
Bachelor’s degree at university	3 (20.0%)
Postgraduate degree (e.g., PGDip)	1 (6.7%)
Masters degree	2 (13.3%)
PhD or Doctorate	1 (6.7%)
Other	2 (13.3%)

Table 9 (continued)

Demographic Characteristics of the Consumer Panel (N = 15)

Demographic variable	Frequency (%)
Employment Status	
Employed full-time	4 (26.7%)
Employed part-time	2 (13.3%)
Employed on a casual basis	0 (0.0%)
Student	2 (13.3%)
Stay-at-home parent	0 (0.0%)
On maternity leave	3 (20.0%)
Unemployed (i.e., looking for work)	0 (0.0%)
Other	4 (26.7%)
Relationship Status	
Single	2 (13.3%)
Married	11 (73.3%)
De facto	1 (6.7%)
Separated	1 (6.7%)
Divorced	0 (0.0%)
Widowed	0 (0.0%)

Expression of Disordered Eating in Pregnancy

Overall, 48 of the 69 potential symptoms reached the consensus agreement criterion across both panels, including behavioural (22 of 27), physical (3 of 14), cognitive (13 of 16), and affective (10 of 12) symptomatology. An additional 20 items were generated and rated only by the consumer panel, with 19 reaching consensus. The item frequency ratings are detailed in Table 10 (full item descriptions can be found in Appendix B). For the purpose of brevity, only the item statistics for the final rating round are reported. Both panels endorsed a similar number of behavioural, cognitive, and affective symptom attributes; however, the professional panel endorsed a greater number of physical symptom attributes compared to the consumer panel (10 vs 3, respectively).

Table 10

Panel Ratings for the Potential Symptom Attributes of Disordered Eating in Pregnancy

	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Behavioural symptom items					
Dietary consumption that does not support a healthy pregnancy	P	4.88 (.33)	5.00	100%	Yes
	C	4.67 (1.05)	5.00	93.3%	Yes
Dieting behaviours (e.g., calorie counting)	P	4.15 (.68)	4.00	92.3%	Yes
	C	4.13 (1.41)	5.00	80.0%	Yes
Inflexibility and rigidity with diet (i.e., strict consumption of diet foods only)	P	4.88 (.33)	5.00	100%	Yes
	C	4.07 (1.03)	4.00	86.7%	Yes
Fasting and/or skipping meals	P	4.88 (.33)	5.00	100%	Yes
	C	4.53 (1.06)	5.00	93.3%	Yes
Use of meal replacements (when not advised by a health professional)	P	4.54 (.81)	5.00	88.5%	Yes
	C	4.40 (1.40)	5.00	86.7%	Yes
Repeated weighing	P	3.85 (.78)	4.00	76.9%	Yes
	C	4.67 (1.05)	5.00	93.3%	Yes
Refusing to eat outside of one's home	P	4.65 (.56)	5.00	96.2%	Yes
	C	4.33 (1.23)	5.00	86.7%	Yes
Eating in secret	P	4.73 (.45)	5.00	100%	Yes
	C	4.60 (1.06)	5.00	93.3%	Yes
Eating an objectively large amount of food	P	3.85 (.54)	4.00	76.9%	Yes
	C	3.93 (1.03)	4.00	80.0%	Yes
Eating for "two"	P	2.46 (.76)	2.00	7.7%	No
	C	3.33 (.72)	3.00	33.3%	No
Eating when not physically hungry	P	3.08 (.56)	3.00	19.2%	No
	C	4.13 (.52)	4.00	93.3%	Yes
Using food to cope with/soothe strong emotions, or reward oneself	P	3.92 (.63)	4.00	84.6%	Yes
	C	4.07 (1.10)	4.00	80.0%	Yes
Eating rapidly and until uncomfortably full	P	4.31 (.62)	4.00	92.3%	Yes
	C	4.13 (1.13)	5.00	80.0%	Yes
Self-induced vomiting	P	4.85 (.46)	5.00	96.2%	Yes
	C	4.53 (1.13)	5.00	86.7%	Yes
Obsessively exercising for the purpose of controlling weight and shape	P	4.15 (.54)	4.00	92.3%	Yes
	C	4.60 (1.06)	5.00	93.3%	Yes
Exercising against medical recommendations	P	4.92 (.27)	5.00	100%	Yes
	C	4.60 (1.06)	5.00	93.3%	Yes
Exercising in secret	P	4.88 (.33)	5.00	100%	Yes
	C	4.80 (.56)	5.00	93.3%	Yes
Refusing to purchase maternity clothing	P	2.96 (.82)	3.00	15.4%	No
	C	2.93 (1.22)	2.00	33.3%	No
Wearing specific clothing to conceal pregnancy	P	2.88 (.71)	3.00	15.4%	No
	C	3.67 (1.05)	4.00	68.8%	No
Misuse of gestational diabetes medication	P	4.96 (.20)	5.00	100%	Yes
	C	4.80 (.78)	5.00	93.3%	Yes

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). † = additional item suggested by professional panel in Round I.

Table 10 (continued)

Panel ratings for the potential symptom attributes of disordered eating in pregnancy.

	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Behavioural symptom items (continued)					
Use of laxatives or enemas to reduce gestational weight gain/induce weight loss	P	4.92 (.27)	5.00	100%	Yes
	C	4.80 (.78)	5.00	93.3%	Yes
Use of appetite suppressants or "diet pills"	P	4.88 (.43)	5.00	96.2%	Yes
	C	4.80 (.78)	5.00	93.3%	Yes
Use of natural supplements (e.g., tea detox)	P	4.81 (.49)	5.00	96.2%	Yes
	C	4.67 (.82)	5.00	93.3%	Yes
Body checking behaviours	P	4.00 (.49)	4.00	88.5%	Yes
	C	4.80 (.41)	5.00	100%	Yes
Self-harm	P	4.85 (.37)	5.00	100%	Yes
	C	4.40 (.91)	5.00	86.7%	Yes
Not consuming enough food during pregnancy to produce milk or sustain breastfeeding, resulting in weight loss and/or binge eating behaviours †	P	4.87 (.34)	5.00	100%	Yes
	C	4.60 (.63)	5.00	93.3%	Yes
Spending an excessive amount of time (i.e., multiple hours per week) researching about the most effective ways to reduce pregnancy weight gain and/or ways to lose weight after birth	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes
Searching for or seeking information about disordered eating in pregnancy	P	–	–	–	–
	C	4.53 (.92)	5.00	86.7%	Yes
Using the pregnancy as a 'valid' excuse/reason to avoid feared foods and/or not violate dietary rules	P	–	–	–	–
	C	4.53 (.52)	5.00	100%	Yes
Obsessively recording anticipated and achieved weight gain and calculating calorie intake and exercise output to ensure only the absolute minimum weight gain (and feeling distressed if anything interferes with this)	P	–	–	–	–
	C	5.00 (.00)	5.00	100%	Yes
Preferring to ensure the nausea and ignore physical hunger signals due to fear of weight gain or changes to shape	P	–	–	–	–
	C	5.00 (.00)	5.00	100%	Yes
Going to bed hungry at the end of the day and thinking about food, but not allowing oneself to eat to subside this hunger	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes
Excessively reassuring doctors/midwives that low weight during pregnancy OR lack of weight gain is nothing to be concerned about by reporting vague eating habits (e.g., "I eat heaps")	P	–	–	–	–
	C	4.80 (.78)	5.00	93.3%	Yes
Requesting early discharge from hospital because of the food that might be served and feeling anxious is this early discharge does not or cannot occur	P	–	–	–	–
	C	4.73 (.59)	5.00	93.3%	Yes
Frequent 'fat talk' (i.e., if a pregnant woman talks a lot about how 'fat' she looks or is)	P	–	–	–	–
	C	4.40 (.91)	5.00	86.7%	Yes
Chewing and spitting out large amounts of food, particularly forbidden foods	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). † = additional item suggested by professional panel in Round I.

Table 10 (continued)

Panel ratings for the potential symptom attributes of disordered eating in pregnancy.

	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Physical symptom items					
Low body weight	P	3.96 (.53)	4.00	96.2%	Yes
	C	3.80 (.56)	4.00	73.3%	No
Losing weight while pregnant	P	4.73 (.60)	5.00	92.3%	Yes
	C	3.80 (.68)	4.00	80.0%	Yes
Inadequate gestational weight gain	P	4.77 (.65)	5.00	96.2%	Yes
	C	4.46 (.64)	5.00	93.3%	Yes
Excessive gestational weight gain	P	3.88 (.65)	4.00	80.8%	Yes
	C	3.80 (.68)	4.00	66.7%	No
Rapid gestational weight gain	P	3.92 (.56)	4.00	80.8%	Yes
	C	3.60 (.83)	4.00	66.7%	No
Dizziness and/or fatigue	P	3.54 (.76)	3.00	46.2%	No
	C	2.93 (.80)	3.00	13.3%	No
Feeling nauseated most of the time	P	2.08 (.85)	2.00	7.7%	No
	C	2.67 (1.18)	3.00	20.0%	No
Severe morning sickness that does not stop after the first trimester (hyperemesis gravidarum)	P	4.31 (.84)	5.00	84.6%	Yes
	C	2.00 (1.36)	1.00	20.0%	No
Dehydration	P	4.58 (.58)	5.00	96.2%	Yes
	C	3.27 (.80)	3.00	33.3%	No
Abdominal bloating	P	3.04 (.60)	3.00	11.5%	No
	C	2.93 (.80)	3.00	13.3%	No
Gastrointestinal discomfort	P	3.00 (.63)	3.00	19.2%	No
	C	2.47 (.99)	2.00	13.3%	No
Unborn baby is small/underdeveloped for gestational age †	P	3.96 (.48)	4.00	87.0%	Yes
	C	3.47 (.74)	3.00	33.3%	No
Asymmetrical or slow foetal growth †	P	3.96 (.48)	4.00	87.0%	Yes
	C	3.53 (.74)	3.00	40.0%	No
The woman's blood tests show electrolyte imbalances (e.g., low potassium) †	P	4.31 (.84)	5.00	84.6%	Yes
	C	4.13 (.74)	4.00	80.0%	Yes
Cognitive symptom items					
Overvaluation of body shape and weight	P	4.42 (.50)	4.00	100%	Yes
	C	4.93 (.26)	5.00	100%	Yes
Perceptual disturbance (e.g., perceiving self to be overweight for pregnancy stage, when objectively not)	P	4.42 (.50)	4.00	100%	Yes
	C	4.87 (.35)	5.00	100%	Yes
Poor body image	P	4.12 (.52)	4.00	92.3%	Yes
	C	4.47 (.99)	5.00	80.0%	Yes
Low self-esteem	P	3.77 (.65)	4.00	73.0%	No
	C	4.20 (.56)	4.00	93.3%	Yes
Rumination about gestational weight gain	P	4.04 (.53)	4.00	88.5%	Yes
	C	4.87 (.35)	5.00	100%	Yes
Rumination about health of baby	P	3.08 (.63)	3.00	15.4%	No
	C	3.07 (.80)	3.00	20.0%	No

**Note*. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). † = additional item suggested by professional panel in Round I.

Table 10 (continued)

Panel ratings for the potential symptom attributes of disordered eating in pregnancy.

	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Cognitive symptom items (continued)					
Fixation on post-partum weight loss	P	4.12 (.52)	4.00	92.3%	Yes
	C	4.80 (.56)	5.00	93.3%	Yes
Self-critical thoughts and fear of criticism	P	3.31 (.79)	3.00	42.3%	No
	C	4.20 (.56)	4.00	93.3%	Yes
Comparing personal eating habits to others	P	3.77 (.59)	4.00	76.9%	Yes
	C	3.87 (.74)	4.00	80.0%	Yes
Need for pregnancy to be "perfect"	P	3.88 (.71)	4.00	76.9%	Yes
	C	4.20 (.78)	4.00	93.3%	Yes
Desire for baby to be "small" or "petite"	P	4.73 (.53)	5.00	96.2%	Yes
	C	4.20 (1.08)	5.00	80.0%	Yes
Suicidal thoughts/ideation	P	4.62 (.94)	5.00	88.5%	Yes
	C	4.40 (.83)	5.00	80.0%	Yes
Frequent comparison of weight and shape, with pregnant and non-pregnant women †	P	4.74 (.45)	5.00	100%	Yes
	C	4.67 (.62)	5.00	93.3%	Yes
Belief that vomiting will not adversely impact the fetus/baby because "all pregnant women vomit" †	P	4.78 (.52)	5.00	96.0%	Yes
	C	4.60 (.74)	5.00	86.7%	Yes
Obsessive thoughts during pregnancy that relate to food (e.g., fear of food contamination, "clean eating" to avoid pesticides) †	P	4.74 (.45)	5.00	100%	Yes
	C	4.47 (.83)	5.00	93.3%	Yes
Obsessive thoughts regarding health and normality of pregnancy †	P	3.96 (.64)	4.00	87.0%	Yes
	C	4.07 (.85)	4.00	80.0%	Yes
Thoughts during pregnancy about using breastfeeding as a purgatory method and/or prolonging breastfeeding for weight loss	P	–	–	–	–
	C	4.73 (.80)	5.00	93.3%	Yes
Agonising and debating the absolute necessity of every food item consumed and/or bargaining with oneself	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes
Urges and thoughts of wanting to vomit to relieve physical or psychological tension	P	–	–	–	–
	C	2.93 (1.22)	2.00	33.3%	No
Thoughts that one does not deserve to eat, and having to justify food consumption 'for the baby'	P	–	–	–	–
	C	5.00 (.00)	5.00	100%	Yes
Thoughts of wanting to be 'just bump' (i.e., weight gain is only acceptable in 'pregnancy-appropriate' areas such as the stomach, but not the arms/thighs etc)	P	–	–	–	–
	C	4.33 (.82)	4.00, 5.00	93.3%	Yes
Thoughts of returning to a restrictive diet once the baby is no longer dependent on mother's body (e.g., to grow in the womb, for breastfeeding, etc)	P	–	–	–	–
	C	4.60 (.74)	5.00	86.7%	Yes
Preoccupation with diets, weight management information, and the lack of weight gained by other pregnant individuals and/or admiration for how rapidly these individuals 'snap back' to their pre-pregnancy body weight and shape	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes

**Note*. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). † = additional item suggested by professional panel in Round I.

Table 10 (continued)

Panel ratings for the potential symptom attributes of disordered eating in pregnancy.

	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Affective symptom items					
Distress regarding changing shape + fear of fatness	P	4.27 (.45)	4.00	100%	Yes
	C	4.53 (1.06)	5.00	86.7%	Yes
Distress or guilt after eating "unhealthy" or "bad" foods	P	4.19 (.49)	4.00	96.2%	Yes
	C	4.53 (.83)	5.00	93.3%	Yes
Mood disturbance	P	3.92 (.80)	4.00	84.6%	Yes
	C	3.13 (.99)	3.00	33.3%	No
Anxiety about certain foods/food groups	P	4.08 (.56)	4.00	84.6%	Yes
	C	4.67 (.49)	5.00	100%	Yes
Feeling "out of control" of one's body	P	4.27 (.45)	4.00	100%	Yes
	C	4.60 (.91)	5.00	86.7%	Yes
Feeling a "loss of control" over eating	P	4.77 (.59)	5.00	92.3%	Yes
	C	4.53 (1.06)	5.00	93.3%	Yes
Guilt after eating (any food)	P	4.35 (.49)	4.00	100%	Yes
	C	4.73 (.46)	5.00	100%	Yes
Feelings of shame + disgust about body	P	4.92 (.27)	5.00	100%	Yes
	C	4.80 (.41)	5.00	100%	Yes
Sensitivity to comments regarding weight, shape, or appearance	P	4.04 (.60)	4.00	92.3%	Yes
	C	4.20 (.94)	5.00	80.0%	Yes
Emotional detachment from pregnancy	P	4.46 (.86)	5.00	84.6%	Yes
	C	4.27 (.82)	4.00	80.0%	Yes
Social isolation	P	4.31 (.97)	5.00	84.6%	Yes
	C	4.47 (.74)	5.00	86.7%	Yes
Interpersonal mistrust	P	3.73 (.72)	4.00	76.9%	Yes
	C	3.73 (.88)	4.00	73.3%	No
Feeling relieved or thankful for pregnancy serving as a valid explanation to avoid certain foods or eating very little	P	–	–	–	–
	C	4.87 (.35)	5.00	100%	Yes
Distress in relation to increased appetite during pregnancy	P	–	–	–	–
	C	4.87 (.35)	5.00	100%	Yes
Feeling resentful toward the baby for needing constant food and nutrients to grow in the womb, followed by significant guilt and shame for feeling resentful	P	–	–	–	–
	C	4.60 (1.12)	5.00	93.3%	Yes

**Note*. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). † = additional item suggested by professional panel in Round I.

Expert-derived symptom template. After the full list of symptom attributes were rated, an expert-derived template based on the 48 agreed symptom attributes was developed to assist clinicians understand the signs and symptoms of disordered eating in pregnancy. This template is shown in Table 11 on the following page.

Table 11

Possible Signs and Symptoms of Disordered Eating on Pregnancy Based on the Shared Findings of the Two Delphi Panels

Behavioural presentation/s	Cognitive/psychological presentation/s

• Indications of dietary restriction, such as:
  o Calorie counting and/or fasting
  o Inflexibility and rigidity with diet (i.e., strict consumption of diet food only)
  o Use of meal replacements when not advised by a qualified health professional
  o Calculating calorie intake to ensure minimum weight gain
• Avoidance of meals and/or changes in eating behaviours, such as:
  o Difficulty or refusal eating out
  o Frequently using the pregnancy as a reason to avoid certain foods or food groups
• Indications of binge eating, such as:
  o Eating an objectively large amount of food
  o Eating in secret
  o Eating rapidly and until uncomfortably full
  o Using food to cope with/soothe strong emotions
• Signs of compensatory weight control behaviours, such as:
  o Self-induced vomiting
  o Misuse of certain medications (e.g., diabetes medication)
  o Use of laxative, enemas, appetite suppressants, or 'diet pills'
  o Use of natural weight loss or weight control supplements (e.g., tea detoxes)
  o Excessively exercising (or exercising against medical recommendations)
  o Chewing and spitting out large amounts of foods, particularly 'forbidden' foods
• Evidence of gestational weight and shape preoccupation, such as:
  o Repeated self-weighing
  o Body checking behaviours
  o Strict recording of calorie intake and weight gain
  o Researching or expressing interest in weight control/loss methods
  o Frequent 'fat talk' (i.e., a woman talking a lot about how 'fat' she looks or is during pregnancy)
• Reporting vague eating habits (e.g., "I eat heaps") when questioned by family, friends, and/or health professionals
• Searching for or seeking information about disordered eating in pregnancy
• Evidence or reports of self-harm and/or suicidal behaviour
	• Considering one's pregnancy body shape and weight to be very important in determining self-worth and importance
	• Preoccupation and rules around, gestational weight and shape changes, such as:
	  o Thoughts of wanting to be 'just bump' (i.e., weight gain is only acceptable in 'pregnancy appropriate' areas such as the stomach but not the arms/thighs)
	• Poor body image and evidence of perceptual disturbance (e.g., verbalisations that one is overweight or has gained too much weight, even when weight gain is within the expected range for gestational stage)
	• Evidence of low self-esteem (e.g., self-critical thoughts, fear of criticism)
	• Signs of perfectionism, particularly in relation to the pregnancy, such as:
	  o Need for the pregnancy to be 'perfect'
	  o Obsessive thoughts regarding the health and normality of the pregnancy
	  o Preoccupation with 'clean eating'
	  o Frequent comparisons of eating habits and/or weight and shape to others (whether pregnant or not)
	• Using the pregnancy to rationalise the presence and impact of disordered eating behaviours (e.g., belief that vomiting with not adversely impact the baby because 'all pregnant women vomit')
	• Evidence that one's body image disturbance has extended to the unborn child (e.g., desire for baby to be 'small' or petite')
	• Negative, unusual, or obsessive thoughts about food/eating, such as:
	  o Thoughts one does not deserve to eat and justifying food consumption 'for the baby'
	• Preoccupation with diet and weight management information during pregnancy
	• Fixation on postpartum weight loss planning (e.g., thoughts of returning to a restrictive diet once the baby is no longer dependent on mother's body and/or prolonging breastfeeding for weight loss.
	• Presence of suicidal thoughts/ideation

Table 11 (continued)

Possible Signs and Symptoms of Disordered Eating on Pregnancy Based on the Shared Findings of the Two Delphi Panels

Physical presentation/s	Affective presentation/s
• Significant weight loss, low weight in relation to stage of pregnancy, and/or inadequate gestational weight gain • Significant and/or rapid weight gain, or excessive weight in relation to stage of pregnancy • Severe morning sickness that does not cease after the first trimester • Dehydration and/or electrolyte imbalances • Unborn baby is small/underdeveloped for gestational age • Asymmetrical or slow foetal growth ➢ *Note*. Differential diagnosis of medical and psychological conditions is still required with any physical presentation	• Distress regarding changing weight and shape, and/or fear of 'fatness' • Negative or unusual attitudes toward food and eating, such as: o Anxiety about certain foods or food groups o Guilt after eating (any food) o Feeling a "loss of control" over eating • Dissatisfaction with body shape and weight during pregnancy, as evidenced by: o Reports of feeling "out of control" of one's body o Sensitivity to comments regarding weight, shape, or appearance o Feelings of shame and disgust about pregnancy body • Negative attitudes toward the unborn baby, such as: o Resent or frustration that the baby requires constant food and nutrients to grow (often followed by significant guilt and shame for feeling resentful) o Emotional detachment from the pregnancy (i.e., consistently talking about the pregnancy as if it were not real) • Significant changes in mood and/or anxiety

Note. The presentations in this table are reflective of the symptom attributes that reached consensus across both panels.

Distinguishing Disordered Eating from Pregnancy-Appropriate Symptoms

As shown in Table 12, 27 of the 33 indicators to distinguish symptoms of disordered eating from pregnancy-appropriate symptomatology reached consensus across both panels. The additional item generated the consumer panel also reached the consensus threshold.

Table 12

Panel Ratings for Potential Factors Relevant in Distinguishing Disordered Eating in Pregnancy from Normative Pregnancy Symptomatology

Distinguishing Foci	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Severity of behaviours	P	4.88 (.43)	5.00	96.2%	Yes
	C	4.80 (.41)	5.00	100%	Yes
Severity of cognitions	P	4.88 (.43)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Frequency of behaviours	P	4.85 (.46)	5.00	96.2%	Yes
	C	4.87 (.35)	5.00	100%	Yes
Frequency of cognitions	P	4.85 (.46)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Dietary behaviours in excess to recommended guidelines	P	4.46 (.71)	5.00	88.5%	Yes
	C	4.13 (.64)	4.00	86.7%	Yes
Dietary behaviours in deficit to recommended guidelines	P	4.73 (.60)	5.00	92.3%	Yes
	C	4.33 (.62)	4.00	93.3%	Yes
Exercise behaviours in excess to recommended guidelines	P	4.35 (.75)	5.00	84.6%	Yes
	C	4.33 (.49)	4.00	100%	Yes
Exercise behaviours in deficit to recommended guidelines	P	3.19 (.90)	3.00	34.6%	No
	C	3.33 (1.11)	3.00	40.0%	No
Appropriateness of gestational weight gain	P	3.96 (.45)	4.00	88.5%	Yes
	C	4.20 (.56)	4.00	93.3%	Yes
Health risk or distress to fetus	P	4.88 (.43)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Health risk or distress to mother	P	4.85 (.54)	5.00	92.3%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Distress of (or worry by) family	P	3.92 (.48)	4.00	92.3%	Yes
	C	4.13 (.92)	4.00/5.00	80.0%	Yes
History of pregnancy complications (e.g., miscarriage, premature labour)	P	3.96 (.48)	4.00	84.6%	Yes
	C	4.67 (.72)	5.00	86.7%	Yes
Level of physical impairment or impact	P	4.04 (.66)	4.00	88.5%	Yes
	C	4.93 (.26)	5.00	100%	Yes
Level of psychological impairment or impact (e.g., affective state of mother)	P	4.31 (.66)	4.00	92.3%	Yes
	C	5.00 (.00)	5.00	100%	Yes

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *not important* to 5 = *very important*). Findings represent the final round.

Table 12 (continued)

Panel Ratings for Potential Factors Relevant in Distinguishing Disordered Eating in Pregnancy from Normative Pregnancy Symptomatology

Distinguishing Foci (continued)	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Level of social impairment or impact	P	4.12 (.59)	4.00	88.5%	Yes
	C	4.93 (.26)	5.00	100%	Yes
Level of relational impairment or impact	P	4.04 (.59)	4.00	84.6%	Yes
	C	4.93 (.26)	5.00	100%	Yes
Degree of flexibility with dietary rules	P	4.58 (.58)	5.00	96.2%	Yes
	C	4.47 (.52)	4.00	100%	Yes
Level of insight and/or denial	P	4.81 (.49)	5.00	96.2%	Yes
	C	4.40 (.83)	5.00	93.3%	Yes
Discrepancy between self-reported functioning and medical observations	P	4.81 (.49)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Discrepancy between the woman's report and partner/family reports	P	4.73 (.53)	5.00	96.2%	Yes
	C	4.73 (.46)	5.00	100%	Yes
Available coping strategies (e.g., emotion regulation skills)	P	4.00 (.63)	4.00	88.5%	Yes
	C	4.80 (.41)	5.00	100%	Yes
Available social support	P	4.92 (.69)	4.00	92.3%	Yes
	C	4.73 (.46)	5.00	100%	Yes
History of any psychiatric condition	P	4.08 (.69)	4.00	88.5%	Yes
	C	5.00 (.00)	5.00	100%	Yes
History of an eating disorder	P	4.85 (.46)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
History of subclinical eating disorder symptoms	P	4.85 (.46)	5.00	96.2%	Yes
	C	4.93 (.26)	5.00	100%	Yes
Family history of an eating disorder	P	4.00 (.57)	4.00	92.3%	Yes
	C	4.20 (.56)	4.00	93.3%	Yes
Younger age (< 30 years)	P	2.88 (.59)	3.00	7.7%	No
	C	1.40 (1.06)	1.00	6.7%	No
Older age (> 30 years)	P	2.85 (.54)	3.00	3.8%	No
	C	1.53 (1.25)	1.00	13.3%	No
Ethnicity	P	2.73 (.67)	3.00	0.0%	No
	C	1.60 (1.12)	1.00	6.7%	No
Primigravidity (first pregnancy)	P	2.96 (.44)	3.00	7.7%	No
	C	2.20 (1.52)	1.00	20.0%	No
Multigravidity (subsequent pregnancies)	P	2.88 (.52)	3.00	3.8%	No
	C	2.13 (1.41)	1.00	20.0%	No
Ability to return to "normal eating" and regain feelings of control (w/out being restrictive) after bouts of pregnancy-related appetite changes †	P	4.52 (.47)	5.00	86.9%	Yes
	C	4.73 (.53)	5.00	100%	Yes
Intent behind the behaviour (e.g., restricting one's food intake is only problematic if the intention is to minimise weight gain or lose weight during pregnancy, as opposed to restricting due to nausea)	P	–	–	–	–
	C	4.93 (.26)	5.00	100%	Yes

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *not important* to 5 = *very important*). Findings represent the final round. † = additional item suggested by professional panel in Round I.

Expert-derived list of key distinguishing factors. After the full list of foci items were rated, a key list of quantitative and qualitative factors for clinicians to consider was developed based on the findings across both panels. This list is shown in Table 13 below.

Table 13

Points for Clinicians to Consider when Evaluating Potential Symptoms of Disordered Eating in Pregnancy

- How often is the symptom/s occurring, and with what intensity?
- What is the context and/or intent of the symptom? (e.g., *is a woman's dietary restriction to reduce nausea or minimise gestational weight gain?)*
- Does the symptom deviate from clinical recommendations during pregnancy (e.g., deficits in dietary intake, excess in exercise behaviours)?
- Is the woman's weight in a healthy range relative to pregnancy stage? Could the symptom negatively impact gestational weight gain?
- Is there an actual or anticipated health risk or distress to the mother and/or unborn child?
- Does a woman's family express concern about the symptom/s?
- Does the woman have a history of pregnancy complications (e.g., miscarriage, premature labour)?
- Is the symptom/s causing physical, psychological, social, and/or relational impairment/difficulty for the woman?
- Does the woman have insight into the presence and impact of the symptom/s?
- Is the woman open to addressing the concern?
- Is there a discrepancy between a woman's self-reported functioning and the results of medical tests/observations?
- Is there a discrepancy between a woman's report of functioning and partner/family reports of functioning?
- Does the woman have a history of mental health conditions, particularly eating disorders/disordered eating?
- Is there a history of disordered eating in the woman's family?

Note. The features in this table are reflective of the distinguishing foci that reached consensus across both panel.

Evaluation of a Broad Diagnostic Guideline

In terms of the broad threshold at which behaviours would be considered 'disordered', the most commonly endorsed response by the professional panel was weekly frequency, closely followed by fortnightly and monthly frequency. Over half the consumer panel indicated symptoms would only need to occur once per month to be considered problematic (see Figure 6).

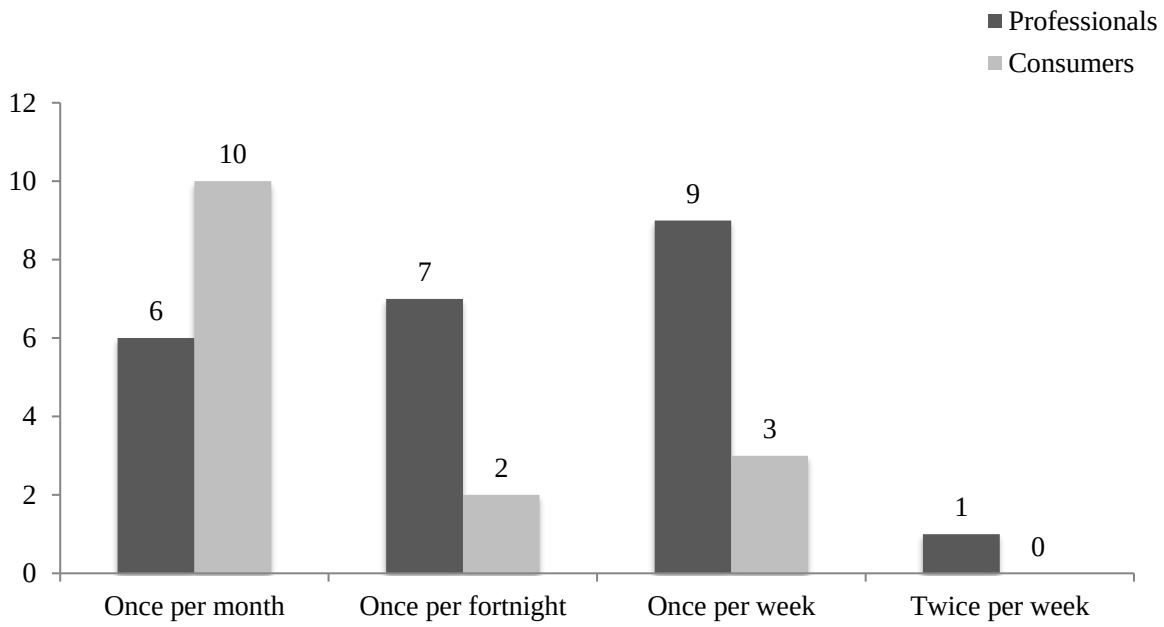

Figure 6. Panel ratings of the most appropriate broad frequency parameter.

Qualitative comments from a professional panel member suggested that including an intensity/severity specifier for the compensatory behaviours would be useful as certain behaviours that could be perceived as compensatory may be medically recommended and are distinct from preventing weight gain in an absolute sense (e.g., weight control strategies for women in the overweight/obese BMI category). Another professional panel member noted certain behaviours may be more or less problematic at certain frequencies and that it is difficult to select a universal frequency for all symptoms. For instance, this panel member noted he/she would have selected a lower frequency option had the words "compensatory

behaviours" been replaced with "self-induced vomiting and/or use of laxatives or diuretics" and a higher frequency option if the compensatory behaviour was dietary restriction.

Qualitative feedback from the consumer panel was more limited. Four panel members commented the presence of any compensatory or binge eating behaviour in pregnancy should be concerning and potentially indicative of an underlying issue. One consumer panel member elaborated, noting that monthly frequency is a reasonable threshold given the distress such symptoms can cause women during pregnancy. Another panel member stated that greater frequency (i.e., fortnightly, weekly) may be reasonable for some symptoms (e.g., dietary restriction, binge eating behaviours), but not for others (e.g., self-induced vomiting, fasting, use of laxatives or diuretics). Such feedback was similar to comments provided by one of the professional panel members. One of the consumer panel members who selected the fortnightly frequency indicated symptom consistency (i.e., at least two occurrences per month) was necessary for such symptoms to be problematic.

Beliefs Regarding Antenatal Assessment of Disordered Eating in Pregnancy

As shown in Table 14, there was clear consensus within and across both panels that assessment of disordered eating should be a routine component of antenatal care for all women, regardless of presenting symptomatology.

Table 14

Panel Ratings for the Implementation of Antenatal Screening

Screening Belief	Panel	Mean (*SD*)	Mode	% of panel agreement	Consensus
Screening for disordered eating should be a routine component of antenatal care (i.e., occur for every woman)	P	4.88 (.43)	5.00	96.2%	Yes
	C	5.00 (.00)	5.00	100%	Yes
Screening for disordered eating in antenatal care should only occur when indicated by presenting signs or historical factors (i.e., selective screening)	P	1.92 (.56)	2.00	3.8%	Yes
	C	1.80 (.78)	2.00	6.7%	Yes
Screening for disordered eating should not occur in antenatal care	P	1.08 (.27)	1.00	0.0%	Yes*
	C	1.13 (.52)	1.00	0.0%	Yes*

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15) Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). Findings represent the final round. *For item 3, 'yes' represents consensus on disagreement.

Qualitative comments echoed this sentiment, with the professional panel noting this was the most important finding to communicate to the scientific community. Five professional panel members commented that, historically, there has been a tendency for health professionals (and society in general) to focus on mental health in the pre-conception and postpartum periods, with less attention to the antenatal period. These panel members highlighted that antenatal mental health needs to be considered to a greater extent, and screening should be a routine and serious component of antenatal care. In general, the professional panel reported brief screening would not be cumbersome for antenatal practitioners and therefore could be easily implemented. One member of the professional panel did, however, express strong concerns across the study that routine screening could generate a high number of "false positives" (i.e., incorrectly identifying healthy pregnant women as potentially having disordered eating) unless the screening instruments had a very strong positive predictive validity. While acknowledging the potential for over identification, other professional panel members noted this trade off would be worthwhile in an area where practitioners often receive minimal training and, therefore, may be less familiar and confident in identifying or screening for presenting symptoms.

Qualitative comments from the consumer panel supported these views, with five panel members indicating that in addition to improving identification, routine screening may raise awareness among pregnant women that such symptoms can occur during pregnancy and that support is available. It was also reported that routine screening could normalise the process of disclosure in this area, potentially de-stigmatising the issue. Selected or indicated-only screening was perceived to 'single out' certain women as being 'disordered', which may negatively affect disclosure and/or the therapeutic relationship. Five consumers cautioned against selective screening as some symptoms of disordered eating may not be visible or noticed during pregnancy and, as such, practitioners may not ask important questions to

garner symptom disclosure. To support this point, two consumers described their personal pregnancy experiences, whereby symptoms of disordered eating went undetected during pregnancy due to non-disclosure and a lack of antenatal screening. Both consumers reported that had discussion of these issues been raised in a routine manner, symptom disclosure would have been likely.

Both consumers and professionals acknowledged that if routine screening was to occur, it should be performed in a way that emphasises the health and wellbeing of the mother and baby, not in a way that focuses on weight, shape, and size, or vilification of body fat. Four professional panel members noted that as pregnancy is characterised by repeated weighing and comments on body shape, weight, and size (both solicited and unsolicited), which can precipitate or exacerbate disordered eating behaviours and cognitions in vulnerable individuals, screening must be performed in a sensitive and caring manner to be of value. There was also a perception, particularly by the professional panel, that routine screening needed to be combined with clinical judgment to guide decisions relating to treatment, whether that be monitoring symptoms, providing education and support, or referral to another health professional (with greater expertise) for more rigorous assessment and care when concerning symptoms are identified.

Evaluation of Potential Assessment Methods

To determine the suitability, practicality, and appropriateness of potential assessment methods, both panels were given a list of 12 methods to evaluate on various 5-point Likert scales. The framing of the question differed slightly between the panels. The professional panel were asked to rate the suitability of the potential methods (1 = *not suitable at all* to 5 = *very suitable*), while the consumer panel was asked to rate the practicality of each method (1 = *not practical at all* to 5 = *very practical*), in addition to a woman's comfort level with each method (1 = *not comfortable at all* to 5 = *very comfortable*). Given that the methods were

rated in a slightly different manner by the two panels, the results are presented in separate tables. Results of the professional panel are presented in Table 15, while results of the consumer panel are presented in Table 17.

Suitability (professionals). As seen in Table 15, the professional panel considered all the potential methods to be suitable for the assessment of disordered eating in antenatal care.

Table 15

Professional Panel Ratings of the Potential Assessment Methods

Screening Method	Panel	Rating	Mean (*SD*)	Mode	% of panel agreement	Cons ensus
Visual observation	P	Suitability	3.62 (.80)	4.00	76.9%	Yes
Physical examination of woman/mother	P	Suitability	4.77 (.51)	5.00	96.2%	Yes
Fetal examination (e.g., ultrasound)	P	Suitability	3.88 (.33)	4.00	88.5%	Yes
Pathology examination	P	Suitability	3.85 (.46)	4.00	80.8%	Yes
Review of medical records	P	Suitability	4.92 (.27)	5.00	100%	Yes
Direct questioning (e.g., *Do you have an eating disorder?*)	P	Suitability	4.92 (.27)	5.00	100%	Yes
Collateral information from support network (e.g., partner, family)	P	Suitability	4.88 (.33)	5.00	100%	Yes
Opportunistic questioning by clinician (unstructured)	P	Suitability	4.81 (.49)	5.00	96.2%	Yes
Brief clinician administered screening (e.g., SCOFF in an oral format)	P	Suitability	4.69 (.88)	5.00	92.3%	Yes
Patient completed screening measures (e.g., SCOFF in a paper-pencil format)	P	Suitability	4.73 (.83)	5.00	96.2%	Yes
Self-report questionnaires (e.g., EDE-Q, EAT, EDI)	P	Suitability	4.85 (.37)	5.00	100%	Yes
Structured clinical interviews (e.g., EDE)	P	Suitability	4.73 (.45)	5.00	100%	Yes

Note. P = professional panel (*N* = 26). C = consumer panel (*N* = 15). Items were rated on a 5-point Likert scale (1 = *not suitable at all* to 5 = *very suitable*). Findings represent the final round.

Qualitative feedback from the professional panel emphasised that brief screening instruments would be ideal for the initial assessment of disordered eating by antenatal providers. Ten panel members noted a brief screening instrument would provide structure or guidance for antenatal providers less experienced with identifying disordered eating, and potentially minimise the risk of comments or questions being misperceived as offensive or

stigmatising by mothers. There was also the suggestion that brief screening instruments may encourage patients to disclose greater clinical information, if delivered in an authentic and caring manner. Such disclosure may be missed when using other screening modalities that do not necessarily open an explicit dialogue between women and clinician (e.g., review of medical records, results of physical tests).

Seven of the professionals noted that although longer questionnaires and structured clinical interviews may entail better psychometric properties and provide greater clinical information, these tools are not feasible in an antenatal setting where practitioners are time limited and ED training may be minimal. Six of the professionals did, however, state that the SCOFF questionnaire (Morgan et al., 1999), an existing screening tool often recommended for use in primary care, might be disadvantageous in pregnancy populations. The two main reasons cited included the poor positive predictive values/sensitivity-specificity levels of the SCOFF demonstrated in non-pregnant populations and the SCOFF items overlapping with pregnancy symptoms.

To explore these concerns in a quantitative manner, the professional panel was asked to rate the suitability of administering the SCOFF in pregnancy in Round III. As shown in Table 16, responses were mixed. Just under half the panel considered the SCOFF to be '*somewhat suitable'* for use during pregnancy, while a third considered the SCOFF to be '*somewhat unsuitable'.* Most of the panel (82.6%) rated items within the SCOFF as overlapping with the experience of pregnancy '*a lot*' (13.0%) or '*a little*' (69.6%).

Table 16

Professional Perceptions of the Suitability of the SCOFF Questionnaire

	VU	SU	N	SS	VS	Mean (*SD*)	Mode
Suitability of the SCOFF instrument	1	9	3	8	2	3.04 (1.15)	2.00

**Note*. SCOFF = Sick, Control, One stone, Fat, Food Questionnaire. VU = very unsuitable, SU = somewhat suitable, N = neither suitable nor unsuitable, SS = somewhat suitable, VS = very suitable. Rated only by the professional panel.

Practicality and acceptability (consumers). In exploring the practicality of the assessment methods, 9 of the 12 potential methods reached consensus in the consumer panel; however, consumers reported they were only comfortable with 7 of these 12 assessment methods (see Table 17). While all the methods perceived to be comfortable were also considered to be practical, there were two approaches the panel appraised as practical but not comfortable: direct questioning by the antenatal care provider and physical examination.

Table 17

Consumer Panel Ratings of the Potential Assessment Methods

Screening Method	Panel	Rating	Mean (*SD*)	Mode	% of panel agreement	Cons ensus
Visual observation	C	Practicality	4.33 (.81)	4.00	93.3%	Yes
	C	Comfort	4.13 (.74)	4.00	93.3%	Yes
Physical examination of woman/mother	C	Practicality	4.67 (.49)	5.00	100%	Yes
	C	Comfort	3.53 (.99)	4.00	60.0%	No
Fetal examination (e.g., ultrasound)	C	Practicality	4.67 (.82)	5.00	93.3%	Yes
	C	Comfort	4.80 (.41)	5.00	100%	Yes
Pathology examination	C	Practicality	4.93 (.26)	5.00	100%	Yes
	C	Comfort	5.00 (.26)	5.00	100%	Yes
Review of medical records	C	Practicality	4.93 (.26)	5.00	100%	Yes
	C	Comfort	4.93 (.26)	5.00	100%	Yes
Direct questioning (e.g., *Do you have an eating disorder?*)	C	Practicality	3.87 (1.13)	4.00	80.0%	Yes
	C	Comfort	2.33 (1.29)	1.00	20.0%	No
Collateral information from support network (e.g., partner, family)	C	Practicality	3.67 (1.05)	4.00	73.3%	Yes
	C	Comfort	2.07 (1.22)	1.00	20.0%	No
Opportunistic questioning by clinician (unstructured)	C	Practicality	3.87 (1.13)	4.00	80.0%	Yes
	C	Comfort	3.80 (.88)	4.00	73.3%	No
Brief clinician administered screening (e.g., SCOFF in an oral format)	C	Practicality	4.67 (.62)	5.00	93.3%	Yes
	C	Comfort	4.33 (1.22)	4.00	93.3%	Yes
Patient completed screening measures (e.g., SCOFF in a paper-pencil format)	C	Practicality	4.73 (.80)	5.00	93.3%	Yes
	C	Comfort	4.80 (.41)	5.00	100%	Yes
Self-report questionnaires (e.g., EDE-Q, EAT, EDI)	C	Practicality	2.93 (1.10)	2.00	26.7%	No
	C	Comfort	3.73 (.88)	4.00	73.3%	No
Structured clinical interviews (e.g., EDE)	C	Practicality	1.40 (1.12)	1.00	6.7%	No
	C	Comfort	1.60 (.99)	1.00	6.7%	No

Note. C = consumer panel (*N* = 15). Items were rated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). Findings represent the final round.

Qualitative feedback from the consumer panel revealed a particular focus on the usefulness of brief screening, with seven consumers commenting that brief screening instruments could be easily incorporated into antenatal care. Eight consumers also noted that brief screening tools would (at a minimum) open a dialogue between a woman and her health professional about the issue of disordered eating in pregnancy, as women may not be voluntarily forthcoming with information due to fear of judgement, criticism, and shame. Six consumers commented that if this conversation was approached in a non-judgmental and sensitive manner this would present an opportunity for women to share their fears/concerns and receive support from their antenatal care provider; however, if the conversation was approached in a stigmatising manner, then this may sever the therapeutic relationship and, in extreme circumstances, result in termination and/or avoidance of care on the mother's behalf.

Discussion

This study used the Delphi consensus technique to garner the collective opinion of two expert panels in relation to disordered eating in pregnancy. Specifically, the current study sought to determine whether consensus could be reached on: 1) the symptomatology of disordered eating in pregnancy, 2) factors that may assist clinicians to distinguish disordered eating from pregnancy-appropriate symptomatology, and 3) whether assessment of disordered eating should be a component of antenatal care and, if so, when assessment should be implemented and what method/s should be used.

Overall, there was a reasonable level of consensus within and across the professional and consumer panels, with four main findings revealed: 1) that while there is a strong degree of similarity between the symptoms associated with disordered eating within and outside of pregnancy, there are certain symptoms that appear to be unique to the pregnant context; 2) the delineation between disordered eating and pregnancy-appropriate abnormal eating is difficult to quantify; however, there are numerous factors and indicators that practitioners can

use when assessing this clinical overlap; 3) antenatal screening for disordered eating is perceived to be crucial and should occur in a routine manner; and 4) implementation of routine screening would be most feasible via use of a brief screening instrument, but this instrument must be pregnancy-specific to be beneficial. Each of these findings is described in further detail and depth below, with potential clinical implications highlighted where appropriate.

Signs and Symptoms Disordered Eating in Pregnancy

In clarifying the manifestation of disordered eating in pregnancy, a range of behavioural, physical, cognitive, and affective elements were identified. There was a strong level of consistency across the panels (47 symptoms meeting consensus across both panels), and generally a high level of consensus (21 with a consensus rate greater than 90% across both panels). Notably, the four symptoms that reached 100 percent consensus across both panels were within cognitive and affective symptom domains. Cognitive symptoms were perceived to be particularly concerning by professionals and consumers given the affective distress these symptoms can produce for a woman. Such distress may have detrimental and lasting impacts on an unborn child, depending on the timing of cortisol exposure (see Davis & Sandman, 2010, for a review). Differences in panel agreement were, however, evident for a subset of symptom attributes. In particular, the professional panel endorsed a greater number of physical symptom attributes than the consumer panel (10 vs. 3, respectively). This difference likely represents the medical knowledge and experiences of the professional panel. As such, it may not have been appropriate to ask the consumer panel to rate such items (Jorm, 2015).

While many of the endorsed symptoms were consistent with those likely observed in a non-pregnant context, several unique pregnancy-specific symptoms were endorsed across both panels including overvaluation of the offspring's weight and shape (e.g., desire for the

baby to be “small” or “petite”), rationalisation of self-induced vomiting as pregnancy-appropriate, and emotional detachment from the pregnancy. Behaviours often normalised outside of pregnancy, such as the use of natural supplements (e.g., tea detoxes) for weight loss, were also considered to be reflective of disordered eating in pregnancy and potentially cause for concern if disclosed to clinicians practicing in this area.

Collectively, the findings suggest that practitioners working with pregnant women should be cognisant of two main factors. First, that an absence of physical or behavioural symptomatology alone does not necessarily imply a woman is unaffected by disordered eating concerns during pregnancy. Previous researchers have also suggested that while observable disordered eating behaviours often reduce during pregnancy, high levels of weight and shape concern, which cannot be easily observed and may not be disclosed freely, often persist (Blais et al., 2000; Crow et al., 2008; Easter et al., 2013; Micali et al., 2007). Second, that disordered eating in pregnancy reflects a spectrum of behaviours that do not necessarily result in physical weight or shape changes, and that particular exploration of binge eating behaviours and cognitions may be justified. This implication supports previous work (Bulik et al., 2007; Knoph Ber et al., 2011; Soares et al., 2009; Watson et al., 2014). Together these findings seem reasonable; yet, antenatal practitioners report a lack of knowledge and confidence in identifying disordered eating symptomatology (Leddy et al., 2009; Morgan, 1999). Furthermore, ED literature suggests that community understanding of the spectrum of disordered eating is poor, with binge eating and/or non-purgatory weight control behaviours often perceived as normative or benign (Mond et al., 2006).

Furthermore, emphasising the finding that disordered eating is multifaceted experience is essential, not only for practitioner awareness in potential screening and detection efforts, but also when educating women who may have limited knowledge or insight in relation to disordered eating symptoms. Historically, presentations of disordered

eating in pregnancy have often been labelled ‘pregorexia’ in popular media, a term describing an excessive fear of pregnancy-related weight gain and engagement in various compensatory behaviours to avoid weight or shape changes that are characteristic of a healthy pregnancy (Hall-Flavin, 2015; Mathieu, 2009; Wallace, 2013). Given the general population is increasingly reliant on popular media sources to obtain important information regarding their health and wellbeing (Fox & Duggan, 2013; Hogue, Doran, & Henry, 2012), it is plausible that women experiencing symptoms inconsistent with the explanation of pregorexia may dismiss or downplay their symptoms. Clinicians interacting with pregnant women must be aware of the potential inaccuracies portrayed in popular media depictions of disordered eating, and the need for appropriate psychoeducation to foster awareness and insight.

Distinguishing the Clinical Overlap Between Pregnancy and Disordered Eating

Arguably one of the most challenging aspects of identifying disordered eating in pregnancy is distinguishing clinical features from the normative pregnancy experience (Easter et al., 2013). While results of the current study do not entirely clarify this nuanced distinction, there was a strong level of agreement across both panels on various quantitative and qualitative factors that might assist practitioners to evaluate concerning symptoms. Some of these factors included: assessment of the woman’s current circumstances (e.g., available coping strategies and/or social support) and historical context (e.g., history of an ED or subclinical disordered eating, history of pregnancy complications); evaluation of familial historical factors (e.g., family history of an ED); and analysis of presenting symptoms (e.g., frequency, severity, and function of symptoms). Practically, information needed to assess these factors could be gathered in routine history taking, followed by more specific questioning, particularly when symptoms are explicit. When symptoms are more subtle or ambiguous, the professional panel noted implementation of clinical judgment would be required. This may include normative comparison of behaviours to clinical guidelines;

evaluation of functional of impairment (e.g., physically, psychologically, socially, interpersonally); and assessment of insight/denial via observed behavioural discrepancies.

It was, however, acknowledged by the professional panel that clinical judgment itself entails a level of ambiguity. As such, a combination of clinical judgment, various qualitative and quantitative indicators, and corroborating information may be required to delineate the clinical overlap. In reality, to comprehensively understand the delineation between disordered eating and pregnancy-appropriate symptomatology, clear guidelines for each individual symptom would need to be developed. This is not only time consuming and resource intensive, but may also be limiting in a setting where fluctuations in symptomatology are common over the course of the pregnancy (Easter et al., 2013). The professional panel noted it is unlikely that a clear cut-off criterion for every symptom is feasible; however, it was acknowledged that practitioners with limited training in identifying disordered eating would potentially benefit from guiding parameters.

When attempting to determine at which frequency symptoms may be considered "disordered" in a broad sense, the professional panelists and consumers both endorsed a less stringent criterion than Easter et al. (2013). In the professional panel, behaviours and/or cognitions were perceived to be of concern if they occurred at least once per week, similar to the frequency parameter adopted by Bulik et al. (2007); however, the consumer panel endorsed more lenient criteria, perceiving symptoms to be concerning if they occurred as little as once per month. While there are acknowledged difficulties with endorsement of simple frequency parameters in terms of lacking contextual and severity information, the frequency difference between the professional and consumer panel potentially indicates that symptoms of relatively low frequency are distressing from a consumer perspective. Further research is required to explore/confirm this finding.

Assessment of Disordered Eating in Pregnancy

Overall, both panels agreed that antenatal screening of disordered eating is crucial and should occur in a routine/universal manner. Several researchers over the past decade have also advocated for such practice (Abraham, 2001; Franko & Spurrell, 2000; Harris, 2010; Leddy et al., 2009; Squires et al., 2014). Recent literature has also suggested that most women perceive mental health screening during pregnancy to be highly beneficial and feel most comfortable when antenatal practitioners initiate the screening process in a routine manner (see Kingston et al., 2015). Although it could be contended that selective screening is more time efficient in already crowded antenatal schedules and consistent with the tailored care approach suggested in the WHO (2016) guidelines for a positive pregnancy experience, there are several arguments against this.

First, selective or indicated screening assumes antenatal providers are familiar with the nosology of EDs, or at least characteristic symptomatology; however, literature has indicated this is often not the case. In Leddy et al. (2009), for example, 88.5 and 96.2 percent of respondents reported a lack of confidence in their ability to identify and manage disordered eating symptoms, respectively. In Morgan (1999), various knowledge deficits were revealed among obstetricians, particularly in relation to the endocrinological and gynecological impact of ED symptoms. Second, although the WHO (2016) guidelines for a positive pregnancy experience report that traditional antenatal care has focused too heavily on clinical services and assessment, it seems unlikely this would relate to perinatal mental health screening given preventative screening in antenatal care has traditionally focused heavily on physical health conditions. Unlike mental health screening which tends to be completed via validated psychometric instruments, biomedical screening is arguably more invasive. While some researchers have proposed the benefits of early detection and intervention do not outweigh the psychological harms of routine screening, most of the

research supporting this view been outside the context of mental health screening. Recent literature has suggested that most women perceive routine mental health screening during pregnancy to have high benefit and low harm (see Kingston et al., 2015, for a full review), with less than four percent refusing or reporting discomfort (Austin et al., 2010; Chew-Graham et al., 2009; Miller et al., 2009). The WHO (2016) guidelines for a positive pregnancy experience have also indicated that women desire additional support coping with psychosocial difficulties during pregnancy and the transition to motherhood. Consistent with this statement, the consumer panel in the current study was highly accepting and supportive of routine antenatal screening of disordered eating. Lastly, although routine screening may be perceived as time consuming and resource intensive, health professionals practicing in settings where routine psychosocial assessment and perinatal mental health screening has been implemented have found this approach to be both feasible and effective (Flynn et al., 2010; Mitchell & Coyne, 2009; Reay et al., 2011; Sword et al., 2008).

Overall, professional panelists considered a range of direct and indirect assessment methods to be relevant when screening for disordered eating in antenatal care; however, this did not necessarily indicate all the methods were practical/feasible for practitioners or comfortable for women. For example, although a structured clinical interview would likely derive robust clinical information, significant practical issues may exist. Implementation of a structured clinical interview would not only be time consuming and cumbersome for antenatal practitioners, particularly on a routine basis, it would also require practitioners to undertake specialised training to administer, score, and interpret the results. Similarly, this method may not be comfortable for women due to the time commitment, in addition to limitations of the structured nature, which may be perceived as interrogative and negatively impact disclosure. Conversely, although brief screening may not garner clinical information as robust as that derived from a structured clinical interview, this method could be easily

integrated and implemented in antenatal care in a variety of formats including clinician administered or patient self-report. Qualitative feedback from the professional panel supported this view, noting that brief screening instruments would be ideal for the initial assessment of disordered eating in antenatal care due to their brevity, flexibility in administration, limited training requirements, and non-threatening nature.

Results of the consumer panel confirmed this, revealing that of the 12 assessment methods perceived to be suitable by the professional panel; only six were comfortable for women, two of which were direct assessment methods including clinician-administered or patient-completed brief screening instruments. The additional four assessment methods were indirect in nature. As such, there appeared to be a common theme across both panels regarding the suitability, feasibility, and appropriateness of clinician-administered or patient-completed brief screening instruments. Direct questioning, unstructured opportunistic questioning, and self-report inventories were not perceived to be comfortable modes of assessment by the consumer advocates.

Consistent with previous literature (Blais et al., 2000; Easter et al., 2013; Koubaa et al., 2005; Patel et al., 2002), concerns regarding the validity of existing screening instruments were expressed by the professional panel, particularly use of the SCOFF questionnaire. The SCOFF questionnaire (Morgan, Reid, & Lacey, 1999) is a brief screening instrument developed as a quick and reliable instrument for non-specialists to identify disordered eating symptomatology and potential EDs. The instrument consists of five key questions (scored in a yes/no format (*no* = 0, *yes* = 1). A score of two or more is generally indicative that deeper and more rigorous assessment is required (Morgan et al., 1999). As the SCOFF was developed for use in non-pregnant populations, validation for use in pregnancy is essential. Although a full discussion of the psychometric properties of the SCOFF is beyond the scope of this chapter, most panelists in the current study expressed concern that items of the SCOFF

overlap with the experience of pregnancy, with one third of the panel suggesting the SCOFF is somewhat unsuitable for use in pregnancy. There was concern this overlap may increase the percentage of false positives (i.e., over-identifying pregnancy symptoms as 'disordered') or, conversely, the rate of false negatives (i.e., under-identifying cases of disordered eating by attributing symptoms to pregnancy). This highlighted a need for systematic exploration of literature to identify existing measures of disordered eating and determine whether such instruments are suitable for use in pregnancy. Results of this systematic review are reported in Chapter 4.

Limitations and Future Research

Although the current study has several strengths to consider, namely the ability to integrate scientific evidence, practitioner experiences, and consumer perspectives to provide a preliminary expert-derived template for understanding and distinguishing disordered eating from pregnancy-appropriate symptomatology, in addition to reaching consensus on the assessment of disordered eating in pregnancy, there are several limitations worth noting, many of which are common to the Delphi methodology.

First, it is acknowledged that the list of symptom attributes and delineating foci generated in the current study is not exhaustive and might have been influenced by the modified Delphi approach, which pre-populates the first round questionnaire following a systematic search. Some researchers have argued the modified Delphi approach may miss key opinions, solutions, or arguments that would be generated via open-ended questions used in the first round of a classical Delphi approach (Hasson et al., 2000). Other researchers, however, have indicated that integrating open-ended questions and comment boxes throughout the first round questionnaire to elicit feedback and suggestions for additional items (as in the current study) can mitigate bias (Green, Jones, Hughes, & Williams, 1999; Trevelyan & Robinson, 2015). Previous research has also suggested the modified Delphi

technique is perceived by panel members to be cooperative and equally as effective as the original technique when executed in a flexible (i.e., allowing panel suggestions/feedback), yet rigorous manner (Eubank et al., 2016; Graefe & Armstrong, 2011; Gustafson et al., 1973). Overall, it is strongly acknowledged that further discussion on the topic of disordered eating in pregnancy is required.

Second, due to the nature of online data collection, it is possible that certain questions may have been misinterpreted; however, care was taken to allow panel members to not only suggest additional items for consideration, but also to provide feedback on the questionnaire process in each section. The aim of this was to ensure any confusion or misunderstandings could be rectified. During Round I, for example, two members of the professional panel expressed difficulty understanding the rating instructions for the distinguishing foci items. This was addressed in Round II by providing additional information and examples in this particular section. This strategy was deemed successful following feedback from the affected panel members. Third, as the Delphi methodology does not allow panelists to discuss topics directly with each other, it is possible that rich information often elicited from intellectual discourse with one's peers may have been missed; however, the anonymity of the panel prevented any power imbalances and group think that may have developed via direct contact. Fourth, although it has been suggested that the Delphi methodology hypothetically prevents power imbalances and group think, some researchers have suggested the process may actually force consensus, as opinions that differ from the accepted norm could be perceived as erroneous (Keeney, Hasson, & McKenna, 2006). Given the high level of consensus displayed in both panels and the relatively strong stability of the consensus items, this seems less likely in the current study.

Fifth, despite efforts to recruit a diverse range of professionals, the panel was mostly comprised of female experts from psychology and psychiatry. Recruiting certain professional

groups, particularly those working in obstetrics and antenatal care, in addition to male experts, was challenging. Possibly the schedules and unpredictable workload of individuals in those fields precluded participation over a six-month period; however, flexible completion options were offered to participants during recruitment. Alternatively, potential panelists from the field of obstetrics may not have identified with the label "expert" due to limited knowledge of disordered eating in pregnancy, which has been revealed in previous research (e.g., Leddy et al., 2009; Morgan, 1991) and may be indicative of a greater educational issue in the field. Further discourse in this area would benefit from a more diverse sample of professionals of both sexes who work directly with disordered eating in an antenatal setting.

Limitations of the consumer panel should also be noted. Although the value of recruiting consumers alongside professionals has been emphasised in recent literature (Jorm, 2015), it is possible the broad criteria for selecting consumers may have affected results, particularly given structured criteria was employed when selecting the professional panel. Expression of interest recruitment in the consumer panel may also have limited the scope of consumers recruited; however, in light of ethical considerations and conceptual issues surrounding the definition of disordered eating in pregnancy prior to the commencement of the study, this approach represented the most suitable recruitment option and had been utilised in previous studies (e.g., Ross et al., 2014). Future research may wish to develop more specific consumer recruitment criteria based on the findings of this study, while also ensuring all viewpoints are considered. A greater number of consumer panel members would also be ideal given the consumer panel size of 15 in the current study was less than the 20 recommended for mental health research (Jorm, 2015); however, other researchers have indicated that panel sizes between 10 and 50 are appropriate and the robustness of a Delphi study is largely dependent on the quality of panel members, not quantity (Linstone & Turoff,

2002; Okoli & Pawloski, 2004). As such, future researchers should carefully consider the balance of quality and quantity when recruiting panel members.

Lastly, the timing discrepancy in administering the Delphi questionnaire rounds between the two panels was undesirable, as this meant new items suggested by the consumer panel at the end of Round I could not be incorporated into the Round II questionnaire for the professional panel. Furthermore, this discrepancy precluded the possibility of evaluating items across both panels during the iterative rounds. As such, the only outcome was to compare the findings of the two independent panels at the end of the study. Future research may benefit from combining consumers and professionals into a single panel (provided questions are appropriate and do not rely on specialist knowledge), or at least ensure concurrent administration of both panels to facilitate feedback and item evaluation across both panels during the Delphi process.

Conclusion

This study represents the first known attempt at utilising the Delphi methodology to achieve consensus among a group of professional experts and consumer advocates in relation to the expression (signs and symptoms), clinical distinction, and assessment of disordered eating in pregnancy. Overall, there was a reasonable level of agreement both within and across the two panels, despite differences in their expertise (Jorm, 2015). Both panels perceived disordered eating to be somewhat similar, yet also distinct, to the experience of disordered eating in a non-pregnancy context. While the exact delineation between disordered eating and pregnancy-appropriate abnormal eating attitudes and behaviours is still not entirely distinguishable, practitioners can attempt to clarify the clinical overlap using a blend of clinical judgment, functional analysis, observed informational discrepancies, assessment of impact and impairment, patient historical factors, and familial historical factors. The importance of routine screening for disordered eating in pregnancy was strongly

emphasised by both panels, highlighting the perceived vulnerability of this powerful biopsychosocial event.

Despite agreement in the professional panel that various assessment methods would be relevant in assessing disordered eating in pregnancy, only a small number of these methods were considered appropriate by the consumer panel. Psychometrically sound brief screening instruments were favoured by both panels, perceived to be most feasible for practitioners to administer and most comfortable for women accessing antenatal care. There were, however, concerns regarding the validity of existing instruments in the pregnancy context. To determine whether existing instruments are appropriate, further exploration in the form of a systematic review is required. If the systematic review in Chapter 4 does not identify a suitable pre-existing measure of disordered eating that could be used in pregnancy, results of Studies 1 and 2 could assist in the development of a pregnancy-specific screening instrument.

CHAPTER FOUR

Disordered Eating Measures Validated in Pregnancy Populations: A Systematic Review

Chapter Overview

This chapter details the methodology and results of the systematic review conducted in the first phase of this research project, following the Delphi studies outlined in Chapter 3. Results of the professional and consumer Delphi studies revealed strong expert consensus that routine screening for disordered eating should be a standard component of antenatal care. The aim of this systematic review was to identify existing measures of disordered eating (conceptualised as including subclinical levels of ED symptoms) and determine whether such instruments are suitable for use in pregnancy or if the development of a pregnancy-specific instrument was required.

Rationale and Objectives

As highlighted in Chapter 1, pregnancy is a sensitive period in which disordered eating symptomatology can arise, be exacerbated, or relapse. The estimated prevalence of disordered eating during pregnancy varies considerably across studies, ranging from 0.6 percent to 27.8 percent, depending on the characteristics of the sample (i.e., pregnancy stage), component of disordered eating being investigated (e.g., cognitive vs. affective), the psychometric instrument employed (e.g., screening tool vs. self-report inventory vs. clinical interview), and the various instrument thresholds used to determine clinically significant scores. Given these varying prevalence estimates, the short- and long- term adverse health consequences for mothers and children, and the unique manifestation of disordered eating during this biopsychosocial event (as outlined in Chapter 3), it is crucial to identify valid and reliable instruments that can be used to measure and discern disordered eating in prenatal care and clinical research. In the current study, disordered eating was conceptualised as

including subclinical levels of ED symptoms, as opposed to alternate forms of disordered eating such external eating, disinhibited eating, or emotional eating.

According to Meades and Ayers (2011), two broad approaches can be undertaken to measure disordered eating symptomatology in the perinatal period: (1) use disordered eating measures developed in other populations and validate them for use with pregnant women; or (2) develop pregnancy-specific measures of disordered eating. To date, research and screening for disordered eating in pregnancy has adopted the former approach, with most researchers using instruments developed and validated in non-pregnant populations, and then suggesting use of these instruments in antenatal care. Examples of these tools include formal self-report inventories such as the Eating Disorders Examination Questionnaire (Astrachan-Fletcher et al., 2008), the Eating Disorder Diagnostic Scale (Easter et al., 2013), the Eating Disorders Inventory-2 (Astrachan-Fletcher et al., 2008), the Eating Attitudes Test (Astrachan-Fletcher et al., 2008; Hawkins & Gottlieb, 2013; Harris, 2010), and the Bulimia Investigatory Test (Harris, 2010). These tools are often used to assist in an ED diagnostic assessment.

Acknowledging the more time-consuming nature of self-report inventories, other researchers have recommended use of brief screening instruments, which typically have 15 items or less and use a simple cut-off score to identify clinical levels of symptomatology, making them ideal for busy clinical settings (Marquer et al., 2012). These instruments do not seek to diagnose a particular condition, rather they aim to identify individuals who may be experiencing concerning symptoms and possibly require further monitoring and/or assessment (Jacobi, Abascal, & Taylor, 2004). Examples of recommended screening instruments include the Sick Control One Fat Food (SCOFF) questionnaire (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; NEDC, 2015) and/or unstructured opportunistic questions (Andersen & Ryan, 2009; Chizawsky & Newton, 2006; Micali, 2010; NEDC, 2015; Ward, 2008).

Extended versions of these informal, opportunistic screening questions, covering both cognitive and behavioural symptomatology, have also been recommended by certain researchers (Chizawsky & Newton, 2006; Ward, 2008). Overall, the SCOFF questionnaire appears to be the most frequent recommendation for detecting disordered eating in pregnancy (NEDC, 2015).

The SCOFF questionnaire is a brief screening instrument, developed by Morgan, Reid, and Lacey (1999). It is promoted as a quick and reliable instrument for non-specialists to identify disordered eating symptomatology and potential EDs. The instrument consists of five key questions (see Table 18) scored in a yes/no format. A score of zero is assigned to *no* responses and score of one is assigned to *yes* responses. Two or more *yes* responses are generally indicative that deeper and more rigorous assessment is required (Morgan et al., 1999). Due to the brevity and accessibility of the instrument, the SCOFF is commonly used in clinical and primary care settings as routine screening for EDs in the general population, and has advantages over traditional, self-report questionnaires which may be more thorough but cumbersome for clinicians to administer (Baudet et al., 2013). The SCOFF can also be administered in either oral or written format.

Table 18

The SCOFF Questionnaire

SCOFF items
1. Do you make yourself **S**ick because you feel uncomfortably full?
2. Do you worry you have lost **C**ontrol over how much you eat?
3. Have you recently lost more than **O**ne stone (6.53kg) in a three-month period?
4. Do you believe yourself to be **F**at when others say you are too thin?
5. Would you say **F**ood dominates your life?

In a recent meta-analytic review (Botella et al., 2013), the SCOFF was found to be highly effective as a brief screening tool for EDs, with pooled estimates from 15 well-controlled studies revealing a sensitivity estimate of 80 percent and a specificity estimate of 93 percent. However, a large proportion of existing validation studies have been completed in specialist ED clinics or student populations, with samples comprising of mostly young (< 40 years), non-pregnant, Caucasian females (Solmi et al., 2015). The homogenous nature of these validation samples potentially limits the generalisability of the SCOFF's accuracy to other populations and circumstances including men, multiethnic populations, older age groups, and women during specific life periods such as pregnancy.

The validity of any instrument, particularly self-report measures, requires re-adjustment for the specific population being examined (Geisinger 1994). Self-report measures developed for use in a particular population may produce flawed or erroneous results when administered in a different population (Myers & Winters, 2002). Data distribution, normative values, and cut-offs may deviate from the original population. As such, self-report instruments must be evaluated in new populations to ensure any variability in measurement is minimised or is similar to the original validation population (Myers & Winters, 2002). It is also accepted that instruments must be validated for use in languages and cultures that differ from the original development and validation sample (Price, 2008). Despite the unique nature of pregnancy, psychological scales developed in non-pregnant populations are frequently used without sufficient evidence to suggest these instruments are suitable or effective (Meades & Ayer, 2011). This could lead to inaccurate interpretations of the data.

One study has recently highlighted the erroneous assumption that psychometric properties remain consistent across all circumstances, with a large multiethnic mixed sex sample (Solmi et al., 2015) revealing a sensitivity value of 53.7 percent and specificity value

of 93.5 percent for the SCOFF questionnaire when benchmarked against the Structured Clinical Interview for DSM-5 conditions (SCID; First, Spitzer, Gibbon, & Williams, 1999). The relatively low sensitivity of the SCOFF in Solmi et al. (2015) indicated a substantial number of non-Caucasian individuals who met criteria for an ED or subthreshold condition were missed during the screening process and, as such, may have gone untreated. While the reasons for the suboptimal performance of the SCOFF in non-Caucasian individuals were not clearly understood, Solmi et al. (2015) suggested the nature of the SCOFF questions may not capture all expressions of ED symptomatology. For example, the emphasis on the fat/thin dichotomy when assessing body dissatisfaction may introduce bias for individuals who are overweight or obese, or where cultural norms differ from what is considered a traditional Western context. The researchers concluded the SCOFF might be suboptimal in screening for EDs in multiethnic communities (Solmi et al., 2015).

These limitations also have direct implications for use of the SCOFF during pregnancy, given that at least four of the questions could be attributed to the experience of pregnancy. For example, the item "Do you make yourself sick because you feel uncomfortably full?" could be attributed to pregnancy 'making' an individual nauseated or sick, while the item "Do you worry you have lost control over how much you eat?" could be ascribed to pregnancy-related appetite increases due to hormonal fluctuations or maternal and/or foetal nutritional needs (King, 2000; Patil, 2012). This overlap may increase the percentage of false positives (i.e., over-identifying pregnancy symptoms as disordered) or, conversely, the rate of false negatives (i.e., under-identifying cases of disordered eating or EDs by attributing symptoms to pregnancy). Both outcomes can lead to suboptimal health care for an individual and the broader community (Buillard & Chiolero, 2015).

Issues with content validity were recently highlighted in a professional guide released by the National Eating Disorders Collaboration (NEDC; 2015), which focused on the

assessment and referral of EDs and disordered eating during pregnancy. Use of the SCOFF was again recommended; however, slight phrasing alterations were suggested and guidelines to assist practitioners distinguish normative pregnancy experiences from disordered eating symptomatology were offered (see Table 19). The impact of these phrasing alterations is, however, unknown, and the tips provided for scoring consideration are only likely to be advantageous when the SCOFF is clinician-administered as opposed to patient-completed.

Table 19

NEDC (2015) Modifications to the SCOFF Questionnaire for Use During Pregnancy

Modified SCOFF items	Consideration points for scoring
1. Do you *forcibly* make yourself **S**ick because you feel uncomfortably full?	• Are there signs of frequent, deliberate vomiting, as opposed to expected levels of morning sickness? • Have signs of vomiting continued despite cessation of morning sickness or nausea?
2. Do you *feel like you lose* **C**ontrol *when you are eating*?	• Is the woman consuming large amounts of food rapidly, or eating an unusually excessive amount of food, even when she is full or not hungry? • Is the woman frequently eating alone?
3. Have you recently lost more than **O**ne stone (6.53kg) in a three-month period?	• Is the mother's weight in the healthy range relative to her stage of pregnancy (i.e., is gestational weight gain adequate?) • Check for signs the woman's weight may be staying the same, despite the additional weight of the baby.
4. Do you believe yourself to be **F**at when others say you are too thin?	• Is there a distortion in the way the woman feels about her body weight or shape? • Does she feel uncomfortable with the way her body is changing due to pregnancy (e.g., "I'm too fat")? • Do others (family, friends) make remarks about her shape or weight (e.g., "You don't seem to be showing much")?
5. Would you say **F**ood dominates your life?	• Is the woman's attitude toward food within the expected responses for pregnancy? • Is she distressed about food or eating? • Has she become overly sensitive or irritable when asked about food or eating?

Note. Italicised text represents wording changes from the original SCOFF tool.

In summary, there is no instrument specifically devised to identify or measure disordered eating in pregnancy. Robust and psychometrically sound measures of disordered eating are needed for screening and research purposes in pregnancy. As such, the aim of this

study was to systematically review the literature to identify and evaluate the performance (outcome) of general measures of disordered eating (intervention) in pregnancy samples (population). Performance was explored by examining the reliability (consistency) and validity (accuracy) of each instrument in pregnancy samples. An additional aim of this review was to identify general measures of disordered eating that demonstrate adequate performance (i.e., are valid and reliable) in pregnancy, and those that require further validation in pregnancy. Performance adequacy was evaluated using a standardised tool.

Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Alessando, Teszlaff, & Altman, 2009) statement was used as a methodological framework.

Data Search

A systematic search of the following electronic databases was undertaken from inception to 30 September 2017: Scopus (Elsevier), Medline (Ovid), PsycINFO (Ovid), Embase (Elsevier), ProQuest Dissertations and Theses, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). The following terms were used to conduct all searches: (eating disorder* OR disordered eating OR inappropriate eating OR maladaptive eating OR problematic eating OR eating disturbance) AND (pregnan* OR antenatal OR perinatal OR intrapartum OR maternity*) AND (screen* OR questionnaire OR scale OR instrument OR measure OR assessment OR tool). The reference lists of the included studies were also crossed checked and relevant citations were manually searched and entered. The primary researcher sought regular expert support from the Faculty Liaison Librarian regarding the search terms, search strategy, and relevant databases.

Eligibility Criteria

Studies were included if they: (1) were published in English; (2) examined the reliability and/or validity of a disordered eating measure (regardless of whether this was the main aim of the study); and (3) the sample or subsample was pregnant at the time of data collection. In instances where different psychometric properties of a measure were assessed by more than one study using the same sample, both studies were included with a note indicating multiple use of a sample. Studies were excluded if the methodological design was inappropriate such as review articles, retrospective studies in the postnatal period, or longitudinal designs that evaluated the psychometric performance of a test from prepartum to postpartum, without any clear distinction between time points. Studies where instruments were administered in a language other than English were also excluded given language alone can impact psychometric performance (Price, 2008). Additional details regarding the main performance outcomes (reliability and validity) are provided below.

Reliability. Reliability is the degree to which a measure/instrument/assessment tool produces stable and consistent results (Furr & Bacharach, 2008). For a scale or instrument to be considered reliable, the scores yielded must represent a proportion of the true state of the assessed (latent) variable (DeVellis, 2012). This implies that the score produced by the scale or instrument should not alter unless there has been an actual change in the variable the instrument is measuring. Thus, an instrument with perfect reliability would reflect the true score of the latent variable only. This seldom occurs as all measurements are influenced by extraneous factors (i.e., error). Accordingly, scale reliability represents the proportion of variance attributable to the true score of the latent variable of interest (DeVellis, 2012). Reliability is a measure of the properties of a test within a particular sample, rather than a property of a test (Vacha-Haase, Kogan, & Thompson, 2000). As such, reliability estimates should be computed in every sample in which the test is utilised.

Several forms of reliability were considered in the current study, including internal consistency, and inter-rater reliability. Internal consistency represents the homogeneity of items within a scale (DeVellis, 2012). According to classical test theory (Feldt & Brennan, 1989; Spearman, 1904), if items in a scale have a strong relationship to the latent construct or variable they are measuring, they will also have a strong relationship to one another. As such, an internally consistent scale demonstrates high inter-item correlations. A minimum Cronbach's alpha coefficient of .70 is recommended for a scale to demonstrate adequate internal consistency for research (Kline, 2000; Nunnally, 1978). Inter-rater reliability represents the degree of agreement or consistency in a rating by two or more observers (DeVellis, 2012). This form of reliability is particularly relevant for clinician-administered and scored instruments such as structured clinical interviews. Several statistical procedures can be used to determine inter-rater agreement. The most common being Cohen's kappa or intra-class correlation coefficients, with values .80 or greater considered minimally acceptable for an instrument (Cichetti, 1994; McHugh, 2012). Test retest reliability, the ability of an instrument to yield consistent scores from one occasion to another (DeVellis, 2012), was not a focus of this systematic review as symptoms of disordered eating are likely to fluctuate in intensity and frequency throughout the course of pregnancy (Easter, 2015). Overall, the reliability of test scores is viewed as a necessary condition to establish validity (Thompson, 2003).

Validity. Validity is the extent to which an instrument measures what it intended to measure in a specific sample (Furr & Bacharach, 2008). More specifically, validity is the degree to which evidence and theory support the interpretation of test scores for the uses proposed by the test developers (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). The validity of an instrument is study specific and must be considered each time the instrument or

measure is selected for a new study, setting, and/or population (Svensson, 2014). Conventional definitions of validity focus on three main forms: content, construct, and criterion-related validity. As such all three were considered in the current study.

Content validity is the extent to which a specific set of items reflects all facets of the construct of interest (Furr & Bacharach, 2008). As such, content validity is strongly linked to the conceptual definition of the construct being examined and the process of developing the items of a scale. Content validity is not usually assessed quantitatively. Instead, qualitative methods such as expert input and feedback during item development are more appropriate. Construct validity is concerned with the internal structure of a test and the way in which the parts of a test relate to each other (Furr & Bacharach, 2008). For an instrument to be a valid measure of a specific construct, the actual structure of the test should match the theoretically based structure of the construct, whether that be uni- or multi- dimensional (Furr & Bacharach, 2008). Factor analytic approaches are mostly commonly used to evaluate the internal structure of a test. Another subtype of construct validity is convergent validity, which takes two measures that are assumed to measure the same underlying construct and demonstrates a reasonable correlation between them. Conversely, discriminant validity is the extent to which an instrument does not correlate with measures that are conceptually different (Furr & Bacharach, 2008). Overall, if an instrument's actual correlations are consistent with the pattern of correlations derived from theoretical expectations, this provides evidence that an instrument is measuring the latent construct of interest.

Criterion-related validity is the extent to which scores yielded from an instrument are correlated with or predict an outcome on another measure or criterion. A criterion can be any variable that an instrument should reasonably correlate with, such as an objective measure (e.g., a true state) or another instrument measuring the same construct (e.g., a gold standard, most commonly a diagnostic interview). When a criterion is measured at the same time as the

index test, criterion validity is known as concurrent validity. When a criterion is measured after the index test has been administered, criterion validity is referred to as predictive validity. In some instances, the criterion may have been administered in a period before the index test, known as postdictive validity. Validity is also concerned with the accuracy of a test or measure. As such, when an index test classifies participants are 'unwell/diseased' or 'well/non-diseased', sensitivity and specificity are often calculated using a criterion (concurrent or predictive) also capable of binary classification. Sensitivity is the ability of an index test to correctly classify an individual as 'unwell/diseased', reflecting the probability of a positive test result or screen when the disease/condition is present (true positive). Conversely, specificity is the ability of an index test to correctly classify an individual as 'well' or 'disease free', reflecting the probability of a negative test result or screen when the disease or condition is absent (true negative). Sensitivity and specificity are inversely proportional. As such, an increase in sensitivity is associated with some decrease in specificity (and vice versa). The relationship between sensitivity and specificity is assessed using receiver operating characteristic (ROC) curves.

Study Selection

The titles and abstracts of all studies identified via the search strategy were screened for inclusion using the eligibility criteria. Studies that did not meet the eligibility criteria from title and abstract screening were removed. If it was unclear whether a study should be included or excluded, the full text was obtained and reviewed according to the eligibility criteria.

Data Quality

The quality of included studies was assessed using a combined checklist based on the quality assessment of diagnostic accuracy studies (QUADAS; Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003) and a checklist developed by Mirza and Jenkins (2004). This

modified checklist has been used in a previous systematic review examining the psychometric properties of anxiety measures in perinatal populations (Meades & Ayers, 2011). A modified checklist was required due to discipline-specific limitations associated with the QUADAS and standards for reporting diagnostic accuracy (STARD) statements (see Streiner, Sass, Meijer, & Furr, 2016). Based on Meades and Ayers (2011), quality criteria in the current study were assessed as present (score of 1) or absent (score of 0) on 11 dimensions: 1) explicit study aims, 2) adequate sample size and/or justification, 3) sample described in sufficient detail, 4) population representative of sample receiving test/measure in practice, 5) inclusion and exclusion criteria clearly stated, 6) use of appropriate reference standard, 7) reliability of measure reported, 8) validity of measure investigated and reported, 9) participant withdrawals and dropouts clearly explained, 10) adequate description of data, and 11) discussion of generalisability. The total number of points received for each study was summed, with quality scores for each study ranging from zero to 11. A score of eight or above was considered reasonable quality.

Performance adequacy was evaluated using standardised criteria developed by Terwee et al. (2007), which considers the quality domains of validity, reliability, and responsiveness. The checklist consists of nine domains (see Table 20 for a description). Across each domain, studies receive either a rating of "+" if positively evaluated, a "?" for an intermediate evaluation, or a "–" for negative evaluation. A "0" is assigned when no information in that domain is available or reported in a study. To assist with the current review (i.e., determine whether general measures of disordered eating, which were developed and standardised in a non-pregnant population, are suitable for use in a pregnancy), the content validity domain was adjusted to include explicit mention of a process whereby an instrument's items were evaluated for appropriateness in a pregnancy context and any modifications were clearly detailed and explained. Similar to Burton et al. (2016), the

reproducibility domain was adjusted to include test-retest correlations greater than .70, with means and standard deviations for both time points reported. This change was adopted to be consistent with the methods more frequently reported in ED literature. Unlike other quality appraisal tools, scores on the Terwee et al. (2007) performance adequacy criteria are not summed into an overall quality score. Terwee et al. (2007) argue that an overall quality score would inaccurately suggest all measurement properties are equally weighted.

Results

Results of the Search Strategy and Study Selection

The literature search yielded 1382 potentially relevant citations. A total of 757 citations remained after removal of all duplicates. These 757 citations were title and abstract screened, with 125 full text articles assessed for eligibility. After assessment of the full-text articles, 8 citations were included and 117 were excluded. The main reason for exclusion of full-text articles was psychometric properties of utilised measures not being explored or reported. In one case (e.g., Crow et al., 2008), inter-rater reliability was reported; however, the longitudinal nature of the study meant reliability estimates were inclusive of pre-partum, intra-partum, and postpartum, rather than pregnancy alone. Another common reason was incorrect study designs (e.g., review articles or retrospective studies in the postnatal period). The PRISMA flowchart of the article selection process can be seen in Figure 7. The data were managed and stored using Covidence (Veritas Health Innovation, 2018), an electronic systematic review platform.

Table 20

Performance Adequacy/Quality Criteria from Terwee et al. (2007)

Property	Definition	Criteria of Adequacy[a,b]
1. Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured in a particular context.	(+) A clear description is provided of the measurement aim, the process for evaluating the suitability of items in a pregnancy context, and any item modifications made are explicitly detailed (?) A clear description of the above mentioned aspects is lacking OR doubtful design or method (–) Description of the above-mentioned elements is lacking and there appears to be no consideration of whether items are appropriate for pregnancy.
2. Internal consistency	The degree which items are intercorrelated, thus measuring the same construct.	(+) Factor analyses are performed on adequate sample (seven times the number of items) AND Cronbach's α (s) between .70 and .95 for each subscale and/or total scale (?) Cronbach's α (s) presented without factor analysis considered OR doubtful design or method (–) Cronbach's α (s) < .70 or > .95 for each subscale and/or total scale (0) No information found on internal consistency
3. Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a "gold standard"	(+) Convincing argument that gold standard is "gold" AND correlation with gold standard is $\geq$ .70 OR AUC is > .70 (?) $\geq$ .70 correlation with gold standard OR AUC > .70 is presented without convincing argument that gold standard is "gold" OR doubtful design or method (–) Correlation with gold standard < .70 (0) No information found on criterion validity
4. Construct validity	The degree to which scores on a questionnaire relate to other measures in a manner consistent with theoretically derived hypotheses concerning the concepts being measured	(+) Explicitly tested for and at least 75% of the results are in the expected direction and size (?) Doubtful design or method (e.g., not explicitly tested) (–) Less than 75% of results as expected (0) No information found on construct validity

Notes. MIC = minimal important changes; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation; AUC = area under the receiver operating characteristic curve; RR = responsiveness ratio; SD = standard deviation

[a] (+) = positive rating; (?) = intermediate or indeterminate rating; (–) = negative rating; (0) = no information available.

[b] Doubtful design or method = lacking clear description of the design or methods of study, sample size smaller than 50 participants (should be at least 50 in every [subgroup] analysis), or any important methodological weaknesses in the design or execution of the study.

Table 20 (continued)

Performance Adequacy/Quality Criteria from Terwee et al. (2007)

Property	Definition	Criteria of Adequacy
5. Reproducibility 5.1 Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	(+) $r > .70$ and means and SD for both time points reported (?) $r > .70$; however, means and SD for both time points not reported (–) $r < .70$ OR doubtful design or method (e.g., time interval not mentioned) (0) No information found on agreement
5.2 Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	(+) *t*-tests, ICC, or weighted $\kappa \geq .70$ (?) Doubtful design or method (e.g., time interval not mentioned, or less valid measures used) (–) *t*-tests, ICC, or weighted $\kappa < .70$ (0) No information found on reliability
6. Responsiveness	The ability of an instrument to detect clinically important changes over time in the construct to be measured	(+) Treatment program outlined and longitudinal expected changes presented and 75% of results are as expected OR SDC < MIC OR MIC outside of LOA or RR > 1.96 or AUC > .70 (?) Doubtful design or method (–) SDC, SDC > MIC, or MIC equals or inside LOA or RR < 1.96 or AUC < .70 (0) No information found on responsiveness
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	(+) < 15% of respondents achieved the highest or lowest possible scores (?) Doubtful design or method (–) > 15% of respondents achieved the highest or lowest possible scores (0) No information found on floor and ceiling effects
8. Interpretability	Degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores.	(+) Mean and SD scores presented for at least four relevant subgroups of patients and MIC defined (?) Doubtful design or method (e.g., data provided on less than four subgroups or MIC not defined) (0) No information found on interpretation

Notes. MIC = minimal important changes; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation; AUC = area under the receiver operating characteristic curve; RR = responsiveness ratio; SD = standard deviation

[a] (+) = positive rating; (?) = intermediate or indeterminate rating; (–) = negative rating; (0) = no information available.

[b] Doubtful design or method = lacking clear description of the design or methods of study, sample size smaller than 50 participants (should be at least 50 in every [subgroup] analysis), or any important methodological weaknesses in the design or execution of the study.

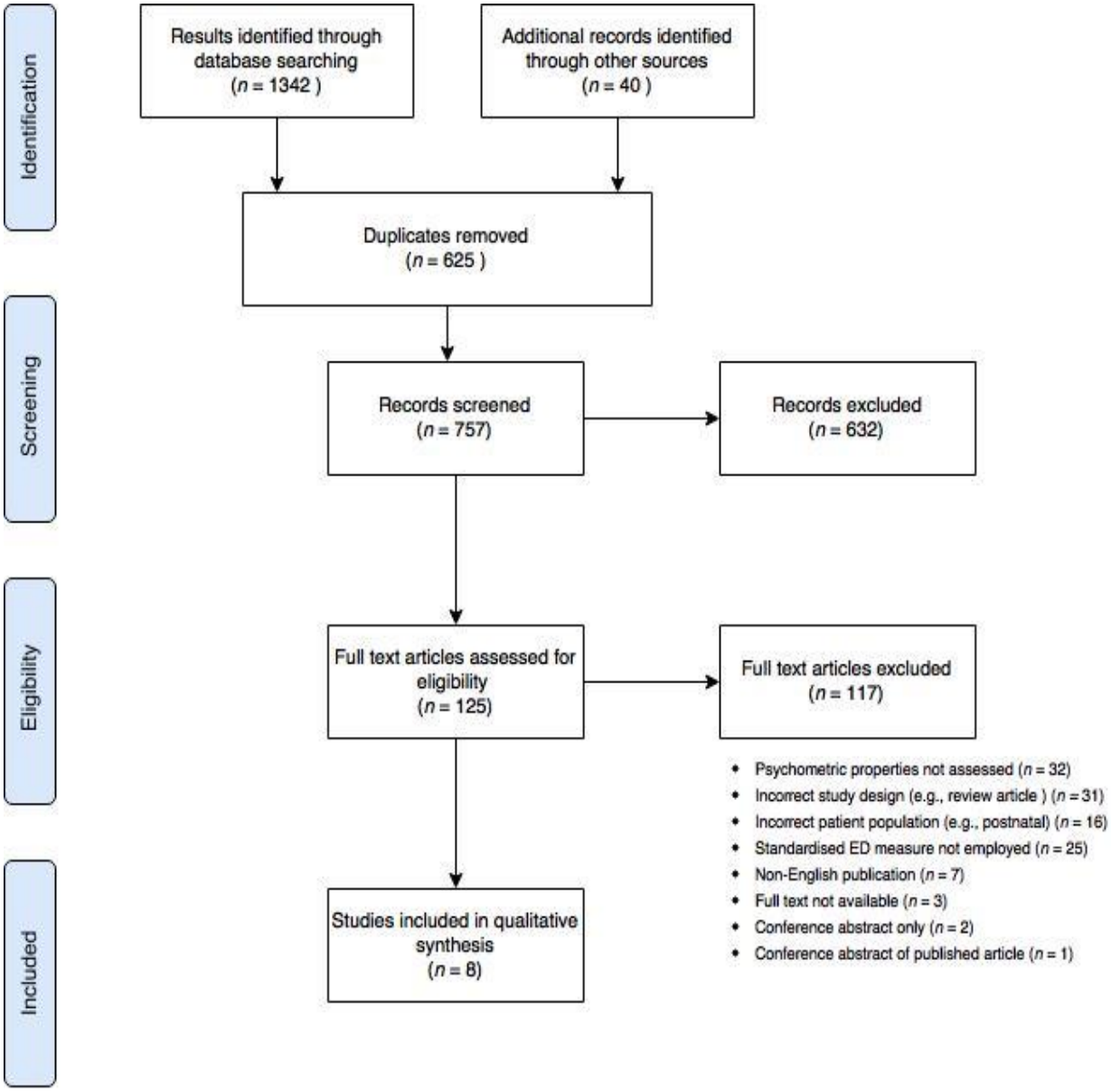

Figure 7. PRISMA article selection flowchart.

Characteristics of Included Studies

Eight publications based on seven studies were included in this systematic review, with 1642 participants. The date range was 2005 to 2017. Country breakdown of the studies was as follows: United States (*n* = 2), Hong Kong (*n* = 1), Sweden (*n* = 1), Pakistan (*n* = 1), Portugal (*n* = 1), and Iran (*n* = 1). All studies utilised cross-sectional designs. Sample sizes ranged from 39 to 426. Ages ranged from 15 to 42 years. Most studies were of good quality

with six (75%) of the eight studies obtaining a score of eight or more on the quality assessment.

Of the eight publications included, only two studies had the aim of assessing the psychometric properties of the employed instruments/measures in a pregnancy sample (Emery et al., 2017; Pettersson et al., 2016). The other six studies reported psychometric information when providing a methodological description of an instrument, as a routine descriptive statistic, and/or indirectly in a correlation coefficient matrix with no explicit explanation. Only two studies assessed validity. One study assessed construct validity (Pettersson et al., 2016), another provided details of discriminant (divergent) validity (Mohamadirizi et al., 2015). No studies assessed criterion-related validity. Seven studies reported reliability, with six reporting internal consistency (Emery et al., 2017; Gonçalves et al., 2015; Lai et al., 2005; Mohamadirizi et al., 2015; Sohail & Muazzam, 2012; Tremblay, 2015). One study reported inter-rater reliability (Kolko et al., 2017). The heterogeneity of the studies and inconsistent psychometric data reported precluded any scope for meta-analysis in this review. Table 21 presents an overview of the included studies.

Psychometric Instruments Identified

Although 16 psychometric instruments that had been used in pregnancy samples were identified during the full-text review process (see Appendix C), only four had psychometric information available, including three different self-report instruments: the Eating Disorders Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994, 2008), the Eating Disorder Inventory-2 (EDI-2; Garner, 1991), and the Disordered Eating Behaviour Scale (DEBS; Muazzam & Khalid, 2011); and one semi-structured clinical interview known as the Eating Disorders Examination (EDE; Cooper & Fairburn, 1987).

Table 21

Overview of Included Studies

Article citation	(#)	Country	*N*	Setting	Sample description (*M* ± *SD*)	Instrument	Validity reported	Reliability reported	Quality score
Lai et al. (2005)	1	Hong Kong	359	Primary care	Age (29.00 ± 4.78) Gestation *(*28.66 ± 9.20) Trimester 10% 1st trimester 20% 2nd trimester 70% 3rd trimester BMI – not reported	EDI-2 (self report)	_	Internal consistency	10
Sohail & Muazzam (2012)	2	Pakistan	300	Primary care	Age (25.78 ± 2.55) Trimester 33.3% 1st trimester 33.3% 2nd trimester 33.3% 3rd trimester BMI – not reported	DEBS (self report)	_	Internal consistency	8

Note. ± = standard deviation; EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination. See Appendix D for a detailed overview of quality scores.

Table 21 (continued)

Overview of Included Studies

Article citation	(#)	Country	*N*	Setting	Sample description (*M* ± *SD*)	Instrument	Validity reported	Reliability reported	Quality score
Mohamadirizi et al. (2015)	3	Iran	213	Primary care	Age (24.12 ± 4.40) Gestation (33.84 ± 3.90) BMI (23.22 ± 4.60)	EDE-Q (self-report)	Discriminant / Divergent	Internal consistency	7
Tremblay (2015)	4	United States	39	Community	Age (26.90 ± 5.12) Gestation *(*28.10 ± 6.51) Trimester 53.8% 2nd trimester 46.2% 3rd trimester BMI – not reported	EDE-Q (self-report)	_	Internal consistency	8
Gonçalves et al. (2015)	5	Portugal	105	Primary care	Age *(M* = 29.95 years) Trimester 100% 3rd trimester BMI (*M* = 28.77)	EDE-Q (self-report)	_	Internal consistency	6

Note. EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination.

Table 21 (continued)

Overview of Included Studies

Article citation	(#)	Country	*N*	Setting	Sample description (*M* ± *SD*)	Instrument	Validity reported	Reliability reported	Quality score
Pettersson et al. (2016)	6	Sweden	426	Primary care	Age (32.50 ± 4.60) Trimester 100% 1st trimester (10-12 weeks)	EDE-Q (self-report)	Construct validity	Internal consistency	8
Emery et al. (2017)	7	United States	129	Community	Age (27.25 ± 5.48) Gestation (14.18 ± 2.15) Trimester 100% 1st trimester (12-20 weeks) BMI (34.38 ± 6.98)	EDE Pregnancy Version (interview)	—	Internal consistency	8
Kolko et al. (2017)	8	United States	200	Same sample as Emery et al. (2017) w additional cases	Age (27.67 ± 5.53) Gestation (15.32 ± 2.40) BMI (34.14 ± 7.23)	EDE Pregnancy Version (interview)	—	Inter-rater reliability	8

Note. ± = standard deviation; EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination. See Appendix D for a detailed overview of quality scores.

Assessment of Psychometric Performance

For each instrument, the psychometric properties reported in the eight publications were assessed using the Terwee et al. (2007) performance appraisal tool. The cumulated results of this evaluation for each instrument are presented in Table 22. Of the four instruments, two did not receive any positive ratings, while two instruments received a positive rating in only one of the nine domains. A description of each measure is detailed following the table, including a summary of the reported psychometric properties.

Table 22

Assessment of Psychometric Performance Using the Terwee et al. (2007) Criteria

	EDE	EDE-Q	EDI-2	DEBS
Content validity	–	–	–	–
Internal consistency	0	+	–	?
Criterion validity	0	0	0	0
Construct validity	0	0	0	0
Reproducibility (agreement)	0	0	0	0
Reproducibility (reliability)	+	0	0	0
Responsiveness	0	0	0	0
Interpretability	0	0	0	0

Note. EDE = Eating Disorder Examination (clinical interview). EDE-Q = Eating Disorder Examination Questionnaire (self report). EDI-2 = Eating Disorder Inventory-2 (self report). DEBS = Disordered Eating Behaviour Scale (self report).

EDE. The EDE is a semi-structured interview that provides a comprehensive assessment of core ED psychopathology. The instrument was developed and validated for use with non-pregnant adults. The EDE consists of 28 items, of which 22 assess core ED symptomatology. These 22 items assess four main areas/subscales: dietary restraint (5 items), eating concern (5 items), weight concern (5 items), and shape concern (8 items) over the previous 28 days. Each item has a number of prompts for the clinician to elicit greater information. As such, the number of questions asked to obtain sufficient information for item scoring is often much higher. A clinician rates the frequency or intensity of each item on 7-

point Likert scales (0 = *feature was absent* to 6 = *feature was present every day* or *to an extreme degree*). Items within each subscale are summed and averaged to provide subscale scores. Summing and averaging the four subscale scores creates a global score. Higher scores are indicative of greater ED-related symptomatology. An additional six items provide information on key behavioural features of EDs in terms of number of episodes of the behaviour and in some instances number of days on which the behaviour has occurred. Responses to the EDE items are commonly mapped to the DSM criteria to determine whether a diagnosis of an ED is present or not. Administration of the EDE takes between 45 and 90 minutes (Fairburn, Cooper, & O'Connor, 2008). Clinicians must be trained in administration of the EDE to ensure a comprehensive understanding of the concepts being addressed and rules governing scoring.

The EDE has strong psychometric properties in non-pregnant adult populations and is widely regarded as the "gold standard" instrument in the assessment and diagnosis of EDs (Berg et al., 2012). Two publications included in the current systematic review (Emery et al., 2017; Kolko et al., 2017) explored the reliability of a pregnancy-modified EDE in a sample of pregnant women who were overweight or obese. Both publications were derived from the same sample. Three major modifications were made to create the EDE pregnancy version (EDE-PV), including a change in the time periods assessed and removal of two items due to a lack of relevance in the context of pregnancy (e.g., loss of menstruation, desire for flat stomach). Emery et al. (2017) revealed the EDE-PV global scale had a less than adequate internal consistency in a pregnancy sample ($\alpha = .65$). Questionable internal consistency estimates were also revealed for three of the four EDE-PV subscales: dietary restraint ($\alpha = .67$), shape concern ($\alpha = .65$), and weight concern ($\alpha = .59$). Due to excessive skewness on the eating concern subscale, Cronbach's alpha was not calculated. Inter-rater reliability of the EDE-PV was found to be high when assessing the intensity and frequency of loss of control

over eating (Kolko et al., 2017). The validity of the EDE in pregnancy was not assessed in any available studies.

EDE-Q. The EDE-Q is a self-report derivative of the EDE interview (EDE; Fairburn & Cooper, 1993), which provides a brief and comprehensive assessment of core ED psychopathology. The instrument was developed in a non-pregnant population. The EDE-Q consists of 28 items, of which 22 assess core ED symptomatology. These 22 items assess four main areas/subscales: dietary restraint (5 items), eating concern (5 items), weight concern (5 items), and shape concern (8 items) over the previous 28 days. The frequency or intensity of each item is rated on 7-point Likert scales (0 = *feature was absent* to 6 = *feature was present every day* or *to an extreme degree*). Items within each subscale are summed and averaged to provide subscale scores. Summing and averaging the four subscale scores creates a global score. Higher scores are indicative of greater ED symptomatology. Various empirically established clinical cut-offs have been used in non-pregnant samples, ranging from $\geq$ 2.30 (Mond et al., 2004b) to $\geq$ 4.00 (Giovazolias et al., 2013; Kelly, Cotter, & Mazzeo, 2012; Penelo et al., 2013). A cut-off of $\geq$ 2.80 has demonstrated optimal sensitivity and specificity in non-pregnant samples (Mond et al., 2008). The EDE-Q is considered a psychometrically sound instrument in a non-pregnant context, demonstrating good reliability and validity across a range of non-pregnant samples (see Berg et al., 2012, for a review).

Three studies in the current systematic review suggested the EDE-Q global scale had excellent internal consistency in pregnancy samples with Cronbach coefficient alphas ranging from .91 (Gonçalves et al., 2015; Mohamadirizi et al., 2015) to .95 (Tremblay, 2015). Internal consistency estimates for the four subscales scores were also strong based on Gonçalves et al. (2015) and Mohamadirizi et al. (2015): weight concern (α = .91), shape concern (α = .89), eating concern (α = .68 to .90), and restraint (α = .82 to .91). Pettersson et al. (2016) examined the factorial validity of the EDE-Q in a pregnancy sample. Results did

not support the theorised four-factor structure of the EDE-Q, favouring a three-factor structure instead.

Results of Pettersson et al. (2016) also suggested that eight of the 22 items were not suited to the perinatal context (inclusive of pregnancy and the postnatal period); however, only three of the eight items suggested for removal had low factor loadings in the antenatal sample (i.e., below .40 on all three factors). Pettersson et al. (2016) recommended the use of a 14-item 'pregnancy optimised' EDE-Q. This optimised EDE-Q reportedly provided a more accurate and reliable measurement of disordered eating symptomatology during pregnancy, compared to the traditional EDE-Q; however, in-depth validation analyses (e.g., criterion-related validity) were not undertaken to determine sensitivity and specificity of the optimised version. No other validity evidence for use of the EDE-Q in pregnancy was found in literature.

EDI-2. The EDI-2 is a revised version of the original EDI, a self-report questionnaire designed to measure psychological and behavioural traits pertaining to EDs, particularly anorexia nervosa and bulimia nervosa. The instrument was developed and validated in a non-pregnant population. The EDI-2 consists of 91-items comprised of three main subscales: drive for thinness (7 items), bulimia (7 items), and body dissatisfaction (9 items). The remaining items contribute to eight additional subscales: ineffectiveness, perfectionism, interpersonal distrust, interoceptive awareness, maturity fears, asceticism, impulse regulation, and social insecurity. Items are rated on a 6-point Likert scale (1 = *never* to 6 = *always*). Ratings in the "non-ED range" (e.g., *never, rarely,* and *sometimes*) are collapsed and given a score of zero. Ratings in the "ED range" (e.g., *often, usually,* and *always*) are given scores of one, two, and three, respectively (Garner, 1991). Higher scores are indicative of a greater tendency to endorse attitudinal and behavioural dimensions pertaining to EDs.

In non-pregnant samples, the EDI-2 is considered a psychometrically sound instrument, with subsequent revisions (e.g., EDI-3) used widely in research and clinical settings as part of a comprehensive diagnostic assessment (see Clausen, Rokkedal, & Rosenvinge, 2009, for a review). Only one study (Lai et al., 2005) in the current systematic review provided estimates of test score reliability for the EDI-2 in a pregnant sample. Results of Lai et al. (2005) revealed good internal consistency for the body dissatisfaction subscale (α = .84), adequate internal consistency for the drive for thinness subscale (α = .72), and poor internal consistency for the bulimia subscale (α = .50). The validity of the EDI-2 in pregnancy was not explored in any available studies.

DEBS. The DEBS is a self-report instrument developed to assess disordered eating behaviour in non-pregnant Pakistani adolescents and adults. According to the test developers, the instrument captures culture-specific disordered eating practices (Muazzam & Khalid, 2011); however, detail of these unique culture specific items is limited. The DEBS consists of 26-items comprising four subscales: social pressures (6 items), eating choices and habits (5 items), eating withdrawal (8 items), and overeating (7 items). Items are scored on a 5-point Likert scale (0 = *never* to 4 = *always*). Higher scores indicate an individual is more prone to engaging in disordered eating beheaviours. Comprehensive exploration of the psychometric properties of the DEBS in non-pregnant populations is limited. Preliminary research by the test developers revealed sound psychometric properties in clinical and community samples of young Pakistani adults (Muazzam & Khalid, 2011). Sohail and Muazzam (2015) recently administered the DEBS to a sample of pregnant women in Pakistan, reporting a Cronbach coefficient alpha of .92 for the full scale. Subscale coefficient alphas were not reported. The validity of the DEBS in pregnancy was not explored in any available studies.

Discussion

This systematic review highlighted the paucity of research validating measures of disordered eating in pregnancy populations. Of the sixteen instruments identified during the full-text review process, only three self-report inventories and one semi-structured clinical interview had some form of psychometric information available. Most studies reported reliability, with only two reporting validity. No studies assessed screening accuracy (i.e., sensitivity and specificity). When the Terwee et al. (2007) criteria were applied, no instrument was able to demonstrate adequate properties in each of the nine domains evaluated. Only two measures obtained a positive rating (one domain each), while the other two measures did not obtain any positive ratings. Recommendations regarding the suitability of each evaluated measure are provided, followed by key issues identified during this review.

Recommendations for use of the EDE-Q, EDI-2, DEBS, and EDE in Pregnancy

Of the four instruments assessed, the EDE-Q had the most psychometric information available; however, there was still insufficient evidence to confer any decision about the appropriateness of the EDE-Q in pregnancy. The EDE-Q demonstrated good to excellent internal consistency at the global and subscale level, but poor factorial validity based on the hypothesised four-factor structure; however, research in non-pregnant samples has also questioned the four-factor structure, suggesting use of the global score is more reliable (Becker et al., 2010; Fairburn et al., 2009). As such, there was preliminary evidence to suggest the global EDE-Q score might be appropriate in pregnancy. No studies assessed criterion-related validity; therefore, the accuracy of the EDE-Q in pregnancy is unknown. Further exploration of the EDE-Q, particularly criterion-related validity, is warranted to determine whether the instrument can be validated for pregnancy and, if so, at which clinical cut-off.

Insufficient evidence precluded a thorough psychometric evaluation of the EDI-2 and DEBS, thereby limiting any recommendations regarding the suitability/appropriateness of these instruments in measuring disordered eating symptomatology in pregnancy. The reported culture specific items of the DEBS are, however, likely to limit generalisability in samples outside Pakistan. Further research exploring the validity of the EDI-2 and the DEBS in pregnancy samples is required to determine their utility in pregnancy.

A similar outcome was revealed for the EDE interview. While the EDE interview is considered the preeminent instrument in the field of EDs and the standard by which all other EDs instruments are validated (Berg et al., 2012), there was no empirical evidence to confirm the EDE is suitable for use in pregnancy. From the two publications reviewed (Emery et al., 2017; Kolko et al., 2017), the EDE was found to have poor internal consistency; suggesting the intended construct may be compromised when used with pregnancy samples. The EDE did demonstrate good inter-rater reliability when assessing loss of control over eating; however, these results should be interpreted cautiously as the sample was comprised of pregnant women in the overweight and obese BMI range only and the modified EDE employed had never been administered prior to the study. Overall, there is insufficient evidence at the current time to suggest the EDE interview can serve as the gold standard instrument for identifying disordered eating in pregnancy. Further research investigating the psychometric properties of the EDE in pregnancy samples is urgently required, particularly exploration of validity if the EDE serves as the gold standard to which existing and new instruments are validated. Without an appropriate gold standard, the development and validation of new self-report instruments is significantly hindered (Greenhalgh, 1997; Troy et al., 1996).

Recommendations for use of the SCOFF in Pregnancy

Despite empirical literature and various antenatal guidelines encouraging clinicians to screen for disordered eating using the SCOFF (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; NEDC, 2015), no published studies were found to support this recommendation. Only one published study (Hubin-Gayte & Squires, 2012) had administered the SCOFF in a pregnant sample, with no psychometric data reported. Although the study was not included in this review, discussion of the results is warranted given widespread recommendations to screen for disordered eating in pregnancy using the SCOFF. In Hubin-Gayte and Squires (2012), the French version of the SCOFF was administered by an obstetrician or midwife at the first antenatal visit (10 to 12 weeks gestation). Of the 285 women who participated (mean age of 32), 188 obtained a negative screen (66%), while 97 obtained a positive screen (34%); however, as the researchers did not verify whether disordered eating was present or not using an appropriate reference standard, the sensitivity and specificity (i.e., accuracy) of the SCOFF in pregnancy remains unclear.

Results of Hubin-Gayte and Squires (2012) also revealed endorsement of a positive response (i.e., indicative of ED symptomatology being present) increased significantly from pre-partum to intra-partum for two of the five SCOFF items, particularly the items relating to loss of control over eating (+10.8%) and intrusive food-related thoughts (+8.9%). Notably, the item pertaining to self-induced vomiting remained relatively stable, despite previous research indicating purging related behaviours often decreased during pregnancy (Easter et al., 2013; Micali et al., 2007). There was also a considerable decrease in affirmative responses to the item querying weight loss of at least six kilograms (–19.5%), as would be expected during pregnancy. While these findings may indicate pregnancy is a high-risk period for disordered eating, it is also likely to reflect the considerable overlap in disordered

eating pathology and pregnancy-related symptomatology, which potentially increased the number of false positives (i.e., incorrectly suggesting disordered eating is present). Conversely, it is also possible symptoms of disordered eating were not detected as participants may have assumed symptoms were benign or pregnancy-appropriate, resulting in a higher level of false negatives (i.e., failing to detect disordered in women who are experiencing such symptoms).

A Need for Pregnancy-Specific Measures of Disordered Eating

A key issue noted in existing literature, and confirmed in this review, is the absence of pregnancy-specific measures of disordered eating. This is in contrast to postnatal depression where several instruments specific to the perinatal period have been developed (e.g., the Edinburgh [Postnatal] Depression Scale and the Postpartum Depression Screen) after researchers acknowledged the poor content validity of general depression measures in pregnancy (Meades & Ayers, 2011). Similar to the measurement of postnatal depression, the present review revealed there is insufficient evidence to support the use of general measures of disordered eating with women who are pregnant. Furthermore, given the overlap between disordered eating and normative pregnancy symptoms, a pregnancy-specific screening instrument that is sensitive to the eating and weight-related changes that occur during pregnancy is warranted. Several researchers have noted the need for such instrument in recent years (e.g., Easter et al., 2013; Pettersson et al., 2016). Terwee et al. (2007) also consider content validity to be one of the most important measurement properties, as it is only if the content validity of an instrument is adequate would one consider using that instrument, and evaluation of other measurement properties would be warranted.

Limitations and Conclusion

The results of this review were limited by the use of stringent performance adequacy criteria defined by Terwee et al. (2007), which has been criticised in literature for being

overly conservative in the allocation of positive ratings (Burton et al., 2016; Reneman, Dijkstra, Geertzen, & Dijkstra, 2010). Despite this, the Terwee et al. (2007) criteria continues to be the most widely utilised tool when evaluating the psychometric performance of instruments (Bird et al., 2012; Cavelti, Kvrgic, Beck, Kossowsky, & Vauth, 2012). The small number of studies available also limited the current review; therefore, the low performance appraisal scores in the current study were mostly due to a lack of available data, rather than poor psychometric performance of the instruments. This issue highlights the dearth of research investigating accurate and reliable screening/measurement of disordered eating symptomatology in pregnancy.

Other than the EDE-Q, which had some preliminary evidence to suggest possible utility, findings of this review revealed little to no evidence to support the use of general measures of disordered eating in pregnancy and strong need for research exploring the validity of existing self-report inventories in pregnancy, including the EDE-Q. Comprehensive validation of these instruments requires validation against an accepted reference standard such as a clinical interview; however, there was insufficient evidence to support the utility of the traditional gold standard instrument (the EDE interview) in pregnancy. Furthermore, despite widespread endorsement, there was also no empirical evidence to support the SCOFF questionnaire as an appropriate screening instrument in antenatal settings. Without reliable and valid measures of disordered eating in pregnancy, researchers and clinicians will have difficulty identifying, measuring, and monitoring disordered eating symptoms in pregnancy. As such, development of pregnancy-specific instruments is required.

CHAPTER FIVE

Development and Evaluation of the Disordered Eating in Pregnancy Scale: A Pregnancy-Specific Disordered Eating Screening Instrument

Chapter Overview

The systematic review in Chapter 4 revealed minimal, if any, evidence to support the use of general measures of disordered eating in pregnancy, highlighting the need for pregnancy-specific measures of disordered eating. As such, the first half of this chapter details the development of a pregnancy-specific disordered eating screening instrument informed by the findings of the Delphi studies outlined in Chapter 3. The second half of this chapter details the psychometric evaluation of the new instrument, in comparison to two well-known, general measures of disordered eating. The prevalence of disordered eating symptomatology in the sample is also reported and reflected upon in comparison to findings from other geographical locations.

Rationale and Objectives

The literature review in Chapter 1 revealed that clinical levels of disordered eating in pregnancy have been linked to several negative consequences, such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and other obstetric and postpartum difficulties (Linna et al., 2014; Watson et al., 2014). Researchers hypothesise such consequences may also extend to subclinical variants (Crow et al., 2008; Harris, 2010); however, comprehensive empirical investigation of such effects is still yet to occur. Current estimates suggest that up to one in four women (25%) may experience symptoms of disordered eating during pregnancy, depending on the psychometric instrument and clinical cut-off employed. Disordered eating symptoms may, however, go undetected and undisclosed due to fear of stigma, poor knowledge and awareness of disordered eating symptomatology during pregnancy, difficulty distinguishing disordered eating from pregnancy-appropriate

symptoms, and the absence of pregnancy-specific assessment instruments (Easter, 2015; Franko & Spurrell, 2000; Franko & Walton, 1993; Freizinger et al., 2010; Hollifield & Hobdy, 1990; Leddy et al., 2009; Morgan, 1997; Newton & Chizawsky, 2006; Tierney et al., 2013).

To minimise the risk of undesirable maternal and infant outcomes and maximise the support provided by antenatal care practitioners, it has been argued that every woman should be screened for disordered eating symptomatology in pregnancy to encourage early identification and management (Abraham, King, & Llewellyn-Jones, 1994; Franko & Spurrell, 2000; Micali & Treasure, 2009). The professionals and consumers in Chapter 3 supported this viewpoint, arguing that pregnancy may represent a period of vulnerability for the precipitation, re-emergence, or exacerbation of disordered eating, and that symptoms may go unnoticed due to the clinical overlap between normative pregnancy experiences and the expression of disordered eating. Early identification of disordered eating can be achieved through the application of an appropriate screening tool (Abraham, King, & Llewellyn-Jones, 1994; Franko & Spurrell, 2000; Micali & Treasure, 2009). That is, a tool that has demonstrated adequate psychometric properties in the population for which it is to be applied.

While several instruments aimed at measuring disordered eating symptoms exist (e.g., EDI, EAT-21, EDDS, BULIT, EDE-Q, SCOFF) and have proven effective in screening for eating disorders in non-clinical populations (Klemchuk, Hutchinson, & Frank, 1990; Mintz & O'Halloran, 2000; Thelen et al., 1991), such instruments were not designed to identify disordered eating symptoms in pregnancy. As the Delphi studies in Chapter 3 revealed, disordered eating in pregnancy entails several unique symptom features. Furthermore, the systematic review in Chapter 4 demonstrated there is little to no evidence to support the use of general measures of disordered eating in pregnancy, other than the EDE-Q, which has a

preliminary evidence base. To facilitate the identification and monitoring of disordered eating symptoms in pregnancy, development of a pregnancy-specific instrument is vital. The development of a psychometrically sound instrument may also encourage a unified research approach when identifying and measuring disordered eating symptomatology in the perinatal context, providing a clearer estimate of disordered eating during this lifespan period.

The Current Study

To address the limitations of existing ED assessment measures, the present study sought to develop and evaluate an instrument that could identify symptoms of disordered eating in pregnancy, conceptualised as subclinical levels of ED symptoms. The desired criteria were that the measure should be brief, inexpensive, easy to administer and understand, and psychometrically sound. Psychometrics is concerned with the theory and technique of measurement in psychology, particularly the objective measurement of a phenomenon and/or construct of interest (Furr & Bacharach, 2008). In 2013, the Journal of Reproductive and Infant Psychology released a consensus statement emphasising the need to test and establish the psychometric properties of instruments used to measure psychological health in the perinatal period (Alderdice et al., 2013). While this clearly applies to the instrument development in this study, the results of the systematic review detailed in Chapter 4 also highlight the relevancy of this statement to other general measures of disordered eating psychopathology such as the EDE-Q and SCOFF, which have limited or no available psychometric evidence. Accordingly, it was hypothesised that:

1. The developed instrument would have a suitable level of internal consistency (i.e., $\geq$ .70; Nunnally, 1978), similar to the EDE-Q, yet much greater than the SCOFF questionnaire.
2. The developed instrument would demonstrate convergent validity with other general measures of disordered eating (e.g., EDE-Q and the SCOFF). Specifically, it was

anticipated the developed instrument would exhibit a high, positive correlation with the EDE-Q, and a moderate, positive correlation with the SCOFF. Such finding was expected given the greater number of items in the EDE-Q and developed instrument, the specific pregnancy-related item content in the developed instrument, and preliminary evidence suggesting the EDE-Q may have possible utility in pregnancy (see Chapter 4).

3. The developed instrument would produce adequate sensitivity ($\geq$ .85) and specificity ($\geq$ .80) estimates using various EDE-Q cut-offs as the criterion, thereby providing evidence of concurrent criterion-related validity. It was also anticipated these estimates would exceed another frequently used screening instrument, the SCOFF questionnaire, when using the same EDE-Q cut-offs as the criterion.
4. The developed instrument would have a high level of scale acceptability and that such ratings would be similar to the general measures of disordered eating included in the study (e.g., EDE-Q and the SCOFF). That is, at least 80 percent of the sample would rate the developed instrument as *comfortable* or *very comfortable* to complete and no significant differences in scale acceptability ratings would be observed across the three measures.
5. The prevalence of disordered eating in the sample, as revealed by the developed instrument, would be similar or slightly higher than estimates reported by previous research (e.g., between 5% and 27.8%).

Method

As the current study involved the development and psychometric evaluation of a new pregnancy-specific disordered eating screening instrument, two phases were required: 1) instrument development and 2) psychometric evaluation. These two phases are detailed below.

Phase I: Instrument Development

To facilitate the instrument development process, a six-step approach derived from overlapping recommendations in DeVellis (2012), Gregory (2015), Kline (2005), and Loewenthal (1996) was employed.

Step 1: Identifying the Construct to Measure

The intention of the current study was to develop an instrument with the capacity to identify possible cases of disordered eating in pregnancy, conceptualised as including subclinical levels of ED symptomatology. In the context of the current study, disordered eating was defined as a range of unhealthy eating and exercise behaviours and cognitions with the potential to negatively impact a woman's emotional, social, and/or physical wellbeing (APA, 2015; NEDC, 2017). While body image is an entwined component or factor in disordered eating, the current instrument was not designed to explicitly or comprehensively measure pregnancy-related body image disturbance, as this construct is broad enough to warrant individual attention in future research. Furthermore, the instrument did not assess external eating (a tendency to eat in response to external cues such as the sight or smell of food) as the Delphi studies in the current project (Chapter 3) did not specifically explore this conceptualisation of disordered eating. The instrument was also not designed to diagnose threshold EDs, rather the aim was to develop a tool to identify women who present with a cluster of subclinical ED symptoms that may warrant further assessment and/or monitoring.

Step 2: Generate an Item Pool

To identify how the construct of disordered eating in pregnancy had been conceptualised and operationally defined in previous research, a comprehensive search of both academic and 'grey' literature was conducted between October and December 2015 (see the Delphi studies in Chapter 3 for additional search details). Overall, 200 sources were used

to develop the preliminary symptoms attributes that were then evaluated and rated in a Delphi study design. Two independent, concurrent panels were employed to obtain subject matter expertise from experienced clinicians and researchers, in addition to women with a lived experience of disordered eating in pregnancy. Both panels reviewed and rated the importance of 60 potential symptom attributes derived from existing literature. An additional seven symptoms were suggested for review following panel feedback. The main objective of this step was to establish consensus on the expression and distinction of disordered eating in pregnancy to inform test development. Following the completion of the Delphi study, the 45 symptom attributes that reached consensus in both panels were used to create an item pool of 42 draft statements, which represented key signs/symptoms to be included in the formal items. These 42 draft statements were then used by the research team to create 19 items, based on the expressed preference for a brief instrument revealed in the Delphi studies. Care was taken to avoid exceptionally lengthy sentences, double-barreled statements, ambiguous wording, and multiple negatives.

Step 3: Determine Measurement and Response Format

To enable rapid assessment of disordered eating in antenatal settings, items were constructed in a forced binary (i.e., yes/no) response format (Loewenthal, 1999). Although a Likert-scale response format was considered, a substantial body of literature has indicated the number of response categories in a scale has little effect on the results obtained (Schutz & Rucker, 1975), temporal stability or inter-rater reliability (Bendig, 1953, 1954; Boote, 1981; Brown, Wilding, & Coulter, 1991; Komorita, 1963; Matell & Jacoby, 1971; Peabody, 1962; Remington, Tyrer, Newson-Smith, & Cicchetti, 1979), and/or concurrent validity (Grassi et al., 2007; Matell & Jacoby, 1971). A significant shortcoming of binary response options, however, is that each item has limited variability. This was considered a reasonable compromise to minimise participant and clinician burden. DeVellis (2012) also suggested

participants are more willing to complete a greater number of binary items, as opposed to items that require concentration on fine category distinction. Thus, sufficient variation in binary scale scores can be achieved by aggregating information over a greater number of items (DeVellis, 2012). Equal weighting was given to each item. Lastly, as previous research has revealed poor disclosure of disordered eating symptoms in pregnancy (Franko & Spurrell, 2000; Franko & Walton, 1993; Hollifield & Hobdy, 1990; Morgan, 1997; Newton & Chizawsky, 2006; Tierney et al., 2013), all items were phrased in first person to reduce the likelihood of unintended, perceived stigma.

Step 4: Consultation with Subject Matter Experts

Once a suitable pool of items had been generated and constructed appropriately, a group of five obstetricians, five midwives, and five clinical psychologists were recruited from two local hospitals to review the item pool for construct relevance; item clarity, conciseness, and phrasing; and potential implementation issues. These experts were recruited based on their clinical experience with the target population and the intended administration setting. This represented the second expert check, as the preliminary item pool was generated from an expert consensus process with researchers, clinicians, and consumers with a lived experience of disordered eating in pregnancy (see Chapter 3).

Feedback from the 15 subject matter experts indicated that 17 of the 19 items were relevant to the construct of interest, presented in a clear and concise manner, appropriately phrased, and non-threatening in nature. There were, however, concerns that two items might be ‘confronting’ for women to answer, particularly if the instrument was clinician-administered. These items related to a woman experiencing disgust in response to her pregnancy body, and a woman feeling resentful toward the unborn child for needing sustained food and nutrients to grow. Although the potential for discomfort was acknowledged, the decision was made to retain these items in their current format to ensure a

clear distinction between normative pregnancy discomfort and disordered eating symptomatology. Minor phrasing changes were suggested and implemented for two items. Additionally, two typographical errors were detected and amended. There was also general feedback that reducing the scale to 15 items or less would make implementation in primary care more feasible.

Step 5: Piloting of Preliminary Item Pool to Refine Items

To identify any structural and/or phrasing difficulties that may have impacted item comprehension and responses, the new instrument was piloted with a sample of 12 pregnant women at various gestational stages (3 in first trimester, 7 in second trimester, and 2 in third trimester). Participants were recruited using convenience and snowball sampling techniques, with online data collection between March and April 2017. To determine the suitability of the items, participants were asked to give qualitative feedback on whether the items were comprehensible and clear (e.g., whether the language, phrasing and terminology used was appropriate), well defined (i.e., did not assess multiple symptoms in an item), and feasible (i.e., level of recall requested was appropriate). The general structure of the scale, specifically level of comfort with the response format, was also assessed

Participants in the pilot study reported the response format was easy to navigate (100%), the level of information requested was reasonable (100%), and almost all the items were phrased clearly and appropriately. Two items were identified to have unclear words or phrases. These words/phrases were therefore modified to aid understanding (e.g., "psychological tension" was reworded to "tension/stress"). It was suggested that two items (e.g., frequent weighing, social comparison of body weight/shape/size) would benefit from frequency specifications to denote problematic behaviour. The broad frequency parameters that reached consensus in the Delphi studies were implemented to facilitate a distinction between normative and non-normative pregnancy experiences. One item was noted to contain

a double-barreled emotive statement (e.g., unhappy and distressed). As such, "unhappy" was removed from the item to ensure a focus on non-normative experiences. Two participants reported it was unclear how two items were distinct from each other (e.g., feeling anxious about eating in general or about eating certain foods versus feeling guilty or distressed after eating). The former item was designed to tap into the thoughts and feelings that occur prior to eating and, as a result, may impact the types of foods women eat (or do not eat) during pregnancy. The latter item related to thoughts and behaviours that occur after food (or a particular food) has been consumed. To ensure this distinction was clear, sequencing words were italicised to indicate the reference period (e.g., before or after eating) and additional detail was added to the latter item for clarity (e.g., "I have felt distressed *after* eating because of its effect on my weight and shape").

The pilot sample commented that one item appeared to assess two symptoms and may benefit from individual items. As such, the item "I spent considerable time researching about the most effective ways to minimise how much weight I will gain during pregnancy and/or how I could rapidly lose weight after I have given birth?" was split into "I spent considerable time researching the most effective ways to minimise how much weight I gain while pregnant*"* and "I spent considerable time researching how I can rapidly lose weight after I have given birth*"*. Three participants noted some items could be answered affirmatively based on medical conditions such as hyperemesis gravidarum. Where possible, the underlying basis of a symptom was emphasised (e.g., to control shape/weight/size). Another participant queried the construct relevance of an item relating to detachment from pregnancy. The decision was made to retain this item for piloting as emotional detachment is reported to be a common experience for women experiencing disordered eating during pregnancy (Astrachan-Fletcher et al., 2008; Park et al., 2003; Patel et al., 2002; Tierney et al., 2013). The researchers were, however, cognisant that detachment from pregnancy can be a

symptom/outcome from other psychiatric conditions or life experiences, rather than disordered eating per se.

Step 6: Penultimate Instrument for Psychometric Evaluation

Following implementation of the recommended and necessary amendments (as described in Steps 4 and 5), a 20-item self-report scale with a binary-response format remained. The next phase of the study was to test and evaluate the performance and psychometric properties of the developed instrument to determine if further scale refinement was required. The utility or appropriateness of the developed instrument also required investigation.

Phase II: Psychometric Evaluation

Participants

A purposive, convenience sample of 508 pregnant women in Australia was obtained using several online recruitment methods (see Procedure). Following the removal of cases with substantially missing data (n = 58) and participants who indicated a multiple birth pregnancy (n = 6), the final sample consisted of 444 women aged 17 to 46 years (M = 29.27, SD = 4.92) carrying a singleton. Most participants were Caucasian (n = 400, 90.1%). More than half of the sample was married (n = 298, 67.1%). Almost three quarters were employed full-time, part-time, or casually (n = 313, 70.4%), while nine percent were on maternity leave (n = 41). Over half the sample reported a university education (n = 246, 55.4%). In terms of obstetric characteristics, 59 percent of the sample was nulliparous (n = 260). Most of the sample reported conceiving without assistance (n = 399, 89.9%), with three quarters noting conception occurred in less than six months (n = 327, 73.6%). Most women were in their second (n = 203, 45.7%) or third (n = 165, 37.2%) trimester. The average gestation was 22.97 weeks (SD = 9.62). See Tables 23 and 24 for additional socio-demographic and obstetric details, respectively.

Table 23

Socio-Demographic Characteristics of Participants (N = 444)

Socio-demographic characteristic	*n*	%
Ethnicity		
Caucasian	400	90.1
Aboriginal or TSI	4	0.9
Asian	26	5.8
Pacific Islander	1	0.2
Hispanic / South American	1	0.2
Middle Eastern	3	0.7
African	1	0.2
Other	8	1.8
Relationship status		
Single	23	5.2
Married	298	67.1
De facto / Engaged	112	25.2
Separated	3	0.7
Divorced	1	0.2
Other	7	1.6
Highest level of education		
High school	68	15.3
Certificate course	58	13.1
Diploma course	71	16.0
Undergraduate degree	133	30.0
Postgraduate degree	113	25.5
Other	1	0.2
Employment status		
Working full-time	209	47.1
Working part-time	73	16.4
Working casually	31	7.0
On maternity leave	41	9.2
Not employed	11	2.5
Homemaker/stay-at-home mum	52	11.7
Student	18	4.1
Other	9	2.0

Note. *n* = number of women in each category. TSI = Torres Strait Islander

Table 24

Obstetric Characteristics of Participants (N = 444)

Obstetric characteristic	*n*	%
Parity		
Nulliparous	260	58.6
Multiparous	184	41.4
Nature of pregnancy		
Planned	278	62.6
Unplanned, but not unexpected	99	22.3
Unplanned and unexpected	63	14.2
Other	4	.9
Time to conception		
Less than 6 months	327	73.6
6 to 12 months	56	12.6
1 to 2 years	35	7.9
2 to 3 years	11	2.5
More than 3 years	15	3.4
Conception nature		
Unassisted	399	89.9
Assisted	45	10.1
Trimester		
First (0-12 weeks)	76	17.1
Second (13-27 weeks)	203	45.7
Third (28-42 weeks)	165	37.2
Pre-pregnancy BMI		
Underweight	14	3.2
Normal weight	219	49.3
Overweight	129	29.0
Obese	82	18.6

Note. *n* = number of women in each category. BMI = body mass index.

Measures

Demographic questions. Participants were asked to supply demographic information to describe the sample. They were also asked to indicate their age, ethnicity, relationship status, highest level of education, and current employment status.

Pregnancy history. To ensure the specified inclusion criteria were met, participants were asked to indicate their current pregnancy status (i.e., pregnant vs. non-pregnant) and whether they were carrying a singleton or multiples (e.g., twins, triplets, etc). Gestational week, parity (i.e., total number of pregnancies), and number of children (not including the current pregnancy) were also requested. To understand the context around the current pregnancy, participants were asked to indicate the length of time to conception (less than 6 months vs. 6 to 12 months vs. 1 to 2 years vs. 2 to 3 years vs. more than 3 years), the nature of the pregnancy (e.g., planned vs. unplanned, but not unexpected vs. unplanned and unexpected vs. other), and the conditions surrounding the pregnancy (e.g., unassisted vs. assisted). Participants were also asked to disclose the type of maternity care being received (e.g., public vs. private vs. other).

Anthropometric questions. To calculate pre-pregnancy and pregnancy body mass indices (BMI), participants were asked to supply their height in centimetres, their current weight in kilograms (pregnancy weight), and their last known weight before conceiving (pre-pregnancy weight). Pre-pregnancy and pregnancy BMIs for each participant were calculated using the following formula: weight (kg) divided by height squared (m^2).

Eating Disorders Examination Questionnaire (EDE-Q; Fairburn & Beglin, 2008). The EDE-Q (Version 6.0) is a 28-item self-report measure of ED symptomatology derived from the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993), a semi-structured interview widely regarded as the instrument of choice for the assessment and diagnosis of DSM eating disorders (Garner, 2002). The EDE-Q provides a similarly comprehensive assessment of specific eating-disordered symptomatology in relatively brief self-report format (approximately 10 to 15 minutes to complete). Numerous studies have confirmed a high level of convergence/agreement between the EDE and EDE-Q when measuring the core characteristics of EDs in community and clinical samples (Anderson, De

Young, & Walker, 2009; Binford, Le Grange, & Jellar, 2005; Carter, Aime, & Mills, 2001; Fairburn & Beglin, 1994; Grilo, Masheb, & Wilson, 2001; Mond, Hay, Rogers, Owen, & Beumont, 2004; Wilfley, Schwartz, Spurrell, & Fairburn, 1997). As such, both instruments are considered 'gold standard' assessment tools in the field of EDs.

In the current study, only the 22 core items were administered. These items assessed four main areas/subscales: dietary restraint (5 items), eating concern (5 items), weight concern (5 items), and shape concern (8 items) over the previous 28 days. The frequency or intensity of each item was rated on various 7-point Likert scales (0 = *feature was absent* to 6 = *feature was present every day* or *to an extreme degree*). Items within each subscale were summed and averaged to provide subscale scores. Summing and averaging the four subscale scores created a global score. Higher scores were indicative of greater ED symptomatology. Three empirically derived global score thresholds (cut-offs) were used as indicator of eating disturbance in the current study: $\geq$ 2.30 (versus < 2.30); $\geq$ 2.50 (versus < 2.50); and $\geq$ 2.80 (versus < 2.80). These cut-offs were selected based on existing support in literature (Mond et al., 2004; Mond et al., 2006; Mond et al., 2008; Pettersson et al., 2016). Furthermore, as there is no established threshold for identifying disordered eating concerns in pregnancy, it was unknown which EDE-Q cut-off would be most suitable.

The EDE-Q has consistently demonstrated good level of internal consistency in clinical and community samples, ranging from .70 to .93 for the four subscales (Aardoom et al., 2013; Bardone-Cone & Agras, 2007; Luce & Crowther, 1999; Mond et al., 2004; Peterson et al., 2007) and .90 to .95 for the global score (Aardoom et al., 2013; Mond et al., 2004; Peterson et al., 2007; Rø, Reas, & Stedal, 2015). Temporal stability has been established in clinical and community samples over a two-week period (r = .66 to .94; Luce & Crowther, 1999; Reas, Grilo, & Masheb, 2006) and five to 14-month period (r = .57 to .79;

Bardone-Cone & Agras, 2007; Mond et al., 2004). Estimates of internal consistency and temporal stability in pregnancy samples are not currently available.

In terms of construct validity, recent studies have not supported the theorised four-factor structure of the EDE-Q (Aardoom et al., 2013; Allen et al., 2011; Barnes et al., 2012; Becker et al., 2012; Darcy et al., 2013; Hrabosky et al., 2008; Peterson et al., 2007; Pettersson et al., 2016; Wade, Byrne, & Bryant-Waugh, 2008). Several studies, including a pregnancy sample (Pettersson et al., 2016), have supported a 3-factor structure (Allen et al., 2011; Darcy et al., 2013; Hrabosky et al., 2008; Peterson et al., 2007). Due to the unclear factor structure, use of the global score only is suggested (Becker et al., 2010; Fairburn et al., 2009). High convergent validity has been established with other measures of ED psychopathology in clinical (Bardone-Cone & Agras, 2007; Celio, Wilfley, Crow, Mitchell, & Walsh, 2004; Grilo, Masheb, & Wilson, 2001) and community (Bardone-Cone & Agras, 2007; Mitsui, Yoshida, & Komaki, 2017) samples. Several studies support the ability of the EDE-Q global score to distinguish between ED cases and non-cases (Aardoom et al., 2013; Machado et al., 2014; Mond et al., 2004; Mond et al., 2008; Wilson et al., 1993) and to detect symptom changes resultant from treatment (Sysko, Walsh, & Fairburn, 2005), providing evidence of strong criterion-related validity and responsiveness. No research has validated the EDE-Q in pregnancy, other than Pettersson et al. (2016), which explored the underlying factor structure of the EDE-Q and suggested removal of eight items for use in pregnancy.

Sick, Control, One stone, Fat, Food (SCOFF; Morgan et al., 1999). The SCOFF questionnaire is a brief screening instrument consisting of five key questions, which can be administered in an oral or written format in less than one minutes (Mond et al., 2008). Each question was scored in a yes/no format, with a score of zero assigned to '*no*' responses and score of one assigned to '*yes*' responses. The number of positive responses was summed, with scores ranging from zero to five. A dichotomous variable of screening-detected

disordered eating was created using the suggested cut-off of scores equal to or greater than two (Hill, Reid, Morgan, & Lacey, 2010; Morgan et al., 1999). Due to the brevity and accessibility of the instrument, the SCOFF is commonly utilised in clinical and primary care settings as routine screening for EDs in the general population, and has advantages over classical, self-report questionnaires which may be thorough but cumbersome for clinicians to administer (Baudet et al., 2013). Numerous researchers have recommended use of the SCOFF to identify symptoms of disordered eating in pregnancy, despite limited psychometric evidence with this population (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2010; NEDC, 2015).

Due to the small number of dichotomous items, Kuder Richardson-20 (KR-20) coefficients have ranged from .43 to .64 (Campo-Arias et al., 2006; Garcia et al., 2011; Hansson et al., 2015; Leung et al., 2015; Mond et al., 2008; Pannocchia et al., 2011). While these coefficients are below the threshold for acceptable internal consistency in a diagnostic test, Streinner (2003) suggests coefficients in this range are acceptable for screening tests. High correlations between item responses on the oral and written SCOFF formats have been revealed, with Kappa coefficients ranging from .85 to .94 (Perry et al., 2002). Intra-class correlation coefficients over a two-week period have ranged from .66 in a sample of 38 high school students in Hong Kong (Leung et al., 2009) to .97 in a sample of 110 previously undiagnosed ED patients (Garcia-Campayo et al., 2005). As such, temporal stability for the SCOFF ranges for moderate to excellent. No studies have explored the reliability of the SCOFF in pregnancy samples.

In terms of validity, most research has supported the proposed unidimensional factor structure (Hansson et al., 2015; Muro-Sans et al., 2008; Pannocchia et al., 2011); however, some studies have revealed a two-factor structure may be appropriate for adolescent girls (Hansson et al., 2015; Muro-Sans et al., 2008). Convergent validity is supported with the

SCOFF displaying moderate to strong point biserial relationships with the EDE-Q total score (r = .60 to .66; Hansson et al., 2015), EDE-Q subscales (r = .39 to .63; Hansson et al., 2015), EDI-2 subscales (r = .53 to .63; Muro-Sans et al., 2008), and EDI-3 subscales (r = .53 to .73; Pannocchia et al., 2015). Using two positive responses as a threshold for potential case identification, sensitivity has ranged from as low as 53.7 percent to as high as 100 percent, while specificity has ranged from 73 percent to 97.1 percent (Baudet et al., 2013; Berger et al., 2011; Caamaño et al., 2002; Cotton et al., 2003; Garcia-Campayo et al., 2005; Garcia et al., 2010; Garcia et al., 2011; Lähteenmäki et al., 2009; Leung et al., 2009; Luck et al., 2002; Mond et al., 2008; Morgan et al., 1999; Muro-Sans et al., 2008; Pannocchia et al., 2011; Parker et al., 2005; Rueda et al., 2005a; Rueda et al., 2005b; Solmi et al., 2015). Further support for the predictive validity of the SCOFF has also been established, with Pannocchia et al. (2011) indicating the SCOFF had a strong ability to distinguish between ED cases, psychiatric controls, and non-clinical controls. To date, no research has validated the SCOFF in a pregnancy sample.

The developed instrument. The penultimate item pool of the developed instrument was also administered. The 20 self-report dichotomous items were scored in a yes/no format, with a score of zero assigned to '*no*' responses and score of one assigned to '*yes*' responses. Administration took less than five minutes. The number of affirmative responses was summed, with initial scores ranging from zero to 20 prior to item refinement. No psychometric information was available prior to the study. The current study aimed to explore the psychometric properties of the developed instrument and determine a suitable cut-off score for potential case identification.

Scale acceptability. Following the completion of each psychometric instrument (e.g., the EDE-Q, the developed instrument, and the SCOFF), participants were asked to indicate how comfortable they felt answering the corresponding scale questions on a 6-point Likert

scale (1 = *very uncomfortable* to 6 = *very comfortable*). As such, three acceptability/comfort ratings were obtained for each participant (one for the EDE-Q, one for the developed instrument, and one for the SCOFF).

Lifetime ED history. Participants were asked to disclose if, at any point in their life, they felt (or had been told by a health professional) they were suffering from an eating disorder or disordered eating. This wording was intended to account for participants who had recovered from an ED, as well as participants that may have struggled with an ED, but never received a formal diagnosis. Participants were able to answer '*yes'*, '*no'*, or '*unsure'*. An open-text box was also provided for participants to contextualise answers, if required.

Procedure

Prior to study commencing, the research was approved by the Bond University Human Research Ethics Committee (#15964). Women who were currently pregnant and above the age of 18 were recruited via several methods including advertisement in the research section of popular parenting websites/forums (e.g., Bubhub, Huggies, Raising Children Network), advertisement in open parenting forums on social media (e.g., Facebook), targeted study promotion using paid Facebook adverts, and advertisement on the Bond University staff and student daily digest. Advertisements in each of these recruitment streams provided an overview of the study, details of participation, contact details for the researchers, and a hyperlink to the online questionnaire. After clicking on the hyperlink, potential participants were directed to a landing page on Qualtrics, a secure online survey platform. The explanatory statement for the study served as the landing page.

Participants were asked to read the explanatory statement, which provided additional study details and noted that involvement in the study was completely voluntary. Informed consent and current pregnancy status were obtained from each participant at the end of the explanatory statement. If a potential participant selected 'no' to either of these questions, she

was directed to the end of the survey and thanked for her time and interest in the study. Participants who were eligible to participate and provided informed consent were then asked to answer a series of questions relating to basic demographic details, pregnancy history, and anthropometric measurements. Participants then completed the EDE-Q, the SCOFF, and the developed instrument. The order of these measures was counterbalanced using the Qualtrics randomiser function to reduce order effects, fatigue effects, and issues with missing data due to attrition. At the bottom of each measure, participants were asked to indicate their level of comfort with the questions asked (i.e., three scale comfort ratings were obtained).

Toward the end of the survey, participants were asked questions pertaining to lifetime ED history. Following completion, participants were given the opportunity to enter a prize draw to win one of ten $50.00 Coles gift vouchers. Interested participants were asked to supply their first name and a contact email address for the prize draw. It was explicitly communicated to participants that these details would not be downloaded with the primary data set to maintain confidentiality. All participants were thanked for completing the survey and given the option to have their responses withdrawn without penalty. Contact details for the researchers and relevant support services were also listed. Completion of the survey took 20 to 30 minutes.

Design

The present study was a cross-sectional correlational design. There were no explicit independent and dependent variables for the main aim of the study (i.e., validation of the developed instrument); however, when examining concurrent criterion-related validity, the cut-off of the developed instrument served as the independent/predictor variable. The three EDE-Q cut-offs ($\geq$ 2.30, $\geq$ 2.50, and $\geq$ 2.80) served as the dependent/outcome variable.

Data Analytic Plan

Part I. To explore the underlying latent structure of the developed instrument and reduce the number of original items, several statistical procedures were performed. Internal consistency was assessed via Kuder-Richardson 20 indices (K-R 20; Kuder & Richardson, 1937) due to the dichotomous nature of the developed instrument. Inter-item and item-total statistics were also inspected to assess the 'contribution' or 'fit' of each item in the scale. The construct validity of the developed instrument was assessed by examining underlying factor structure. Item difficulty was assessed using the mean score of each item of the developed instrument. Discrimination indices for each item of the development instrument were calculated using the extreme group method (Furr & Bacharach, 2008; Kline, 2005) to determine how well each item could discriminate between women with high and low total scores on the developed instrument. The performance of each item of the developed instrument in correctly identifying cases and non-cases at each EDE-Q cut-off ($\geq$ 2.30, $\geq$ 2.50, and $\geq$ 2.80) was evaluated using chi-square analyses. Items were removed or retained following consideration of all item reduction analyses.

Part II. Following the scale refinement / item reduction analyses in Part I, psychometric assessment of the developed instrument in comparison to the EDE-Q and SCOFF was undertaken in Part II. When a new instrument is developed, particularly a screening test, it must be benchmarked against an agreed upon 'gold standard' test (Greenhalgh, 1997); however, gold standards are often not available (Troy et al., 1996), as is the case for the assessment/measurement of disordered eating in pregnancy. As the systematic review in Chapter 4 revealed, the EDE-Q is the closest gold standard proxy at the current time, over and above the EDE interview, which is traditionally used as the gold standard comparison in ED research. As such, the EDE-Q was used as the reference standard for the validation of the developed instrument. As there is no established EDE-Q cut-off for

identifying disordered eating in pregnancy, the developed instrument was benchmarked at three different EDE-Q cut-off points: (i) a global score $\geq$ 2.30, (ii) a global score $\geq$ 2.50, and (iii) a global score $\geq$ 2.80.

The total sample was then divided into two subsamples: Subsample A (primary development sample) and Subsample B (cross-validation sample). The process to create these two subsamples is outlined at the beginning of the Results section. To examine the reliability of the developed instrument (after refinement in Part I), EDE-Q, and SCOFF, internal consistency estimates were calculated for each subsample. K-R 20 indices were calculated for the developed instrument and the SCOFF due to their binary response format. Cronbach's alpha coefficient was calculated for the EDE-Q. The latent structure of the refined developed instrument was re-evaluated using principal components analysis. Convergent validity in each subsample was examined using Pearson product-moment bivariate correlations for the total developed instrument score, the EDE-Q global score, the total SCOFF score.

Criterion-related validity in each subsample was assessed using receiver operating characteristic (ROC) curve analyses. ROC curves are a fundamental tool when evaluating the accuracy of screening and diagnostic tests (Hajian-Tilaki, 2013; Zweig & Campbell, 1993). ROC curves graphically plot the ability of a scale to predict true positives (proportion of people who were correctly classified as positive screen) against the rate of false positives (proportion of people who were incorrectly classified as positive screen). The area under the curve (AUC) is a measure of how well the scale under examination can distinguish between two groups (cases vs. non-cases) according to the different thresholds of the reference criterion. A scale or test with perfect discrimination between the two groups would have an ROC curve that passes through the upper left-hand corner of the ROC graph. As such, the closer the ROC curve is to the upper left-hand corner, the greater the accuracy of the test (Zweig & Campbell, 1993). The area under the ROC curve also measures accuracy, with an

area of .90 to 1.00 representing an excellent test, .80 to .90 representing a good test, .70 to .80 representing a fair test, .60 to 70 representing a poor test, and .50 to .60 representing a test of little to no value (Zweig & Campbell, 1993). Using the three EDE cut-offs as a reference criterion/caseness proxy, six separate ROC analyses were performed (3 for each subsample) to determine the cut-off score of the developed instrument that provided optimal trade-off between sensitivity (proportion of true cases screening positive [Se]) and specificity (proportion of true non-cases screening negative [Sp]). Sensitivity and specificity were considered equally important in the current study to ensure accuracy was not achieved at the cost of over-identification. For comparative purposes, identical ROC analyses were also performed for the SCOFF.

Positive and negative likelihood ratios (LRs) for the developed instrument and SCOFF were also calculated. LRs are another useful measure of accuracy and can help researchers determine the potential utility of a particular test (Deeks & Altman, 2004). LRs represent the probability that a test result is correct, divided by the probability that the test result is incorrect. The sensitivity and specificity of the scale/test are used to generate a LR, which is calculated for both positive and negative test results. As such, LRs are not impacted by the prevalence of a condition in a population, one of the main limitations of positive and negative predictive values (Maxim et al., 2014). A positive likelihood ratio (LR+) indicates how much more likely a positive test result is to occur in participants with the condition compared to those without the condition (Sedgwick, 2011). The further a LR+ is from one, the more accurate a test is. Diagnostic tests typically have an LR+ value greater than 10; however, screening tests are often much lower (Deeks & Altman, 2004). A negative likelihood ratio (LR-) indicates how much more likely a negative test result is to occur in participants with the condition compared to those without the condition (Sedgwick, 2011). LR- values should be less than one because those with the condition should not be obtaining

a negative test result. Good diagnostic tests typically have LR- values less than 0.1; however, this value may be slightly higher in screening tests (Deeks & Altman, 2004). Mann-Whitney U tests were also employed to examine mean score differences between cases and non-cases on each individual item.

Part III. After a suitable cut-off for the developed instrument was identified, participants in each subsample were coded accordingly (i.e., 0 = negative screen and 1 = positive screen). Basic frequency analyses were then used to determine the prevalence of disordered eating according to the developed instrument, the EDE-Q (at all three cut-off points), and the SCOFF (at the specified cut-off point of ≥ 2). Demographic and obstetric differences between cases and non-cases were explored using independent-samples t-tests (continuous variables) or chi-square tests (categorical variables).

Results

The data were analysed using SPSS Version 24. An alpha level of .05 was utilised to determine the statistical significance of all results, unless stated otherwise. Initially, the data were screened for coding errors and missing values. Visual inspection identified 111 cases where individuals had clicked on the survey link but had not commenced the study ($n = 47$) or there was substantially missing data ($n = 64$). The remaining 444 participants consented to the study, met the inclusion criteria, and had complete data.

To counteract shrinkage in validity estimates, the decision was made to split the total sample into two subsamples (DeVellis, 2012). Using the random function in Microsoft Excel, a list of random numbers was generated for 444 cases. These random numbers were then inserted as a variable in the SPSS data file (i.e., each data point was assigned a random number). The random number variable was then sorted in ascending order. The first 222 cases were grouped into Subsample A (primary development sample). The second 222 cases were grouped into Subsample B (cross-validation sample). The main function of Subsample

A was to evaluate the items of the developed instrument using appropriate item statistics, calculate alphas, and determine an optimal cut-off score that maximises sensitivity and specificity (DeVellis, 2012). Subsample B was then used to replicate these findings.

Part I – Scale Refinement / Item Reduction Analyses

Inter-item correlations. Spearman's correlations were performed to examine the extent to which scores on one item were related to scores on all other items in the scale. Inter-item correlations reveal the extent to which items on a scale are measuring the same construct/content (Cohen & Swerdlik, 2005), providing a measure of homogeneity, unidimensionality, and item redundancy (Briggs & Cheek, 1986; Cortina, 1993; Green, Lissitz, & Mulaik, 1977; Piedmont, 2014). Previous researchers have suggested that for a set of items, the average inter-item correlation falls between .20 and .40 (Briggs & Cheek, 1986; Clark & Watson, 1995; Piedmont, 2014). An average correlation within this range suggests that while the items are homogenous, they possess a sufficient level of unique variance to avoid being isomorphic with each other (Piedmont, 2014).

For the new screening instrument, the average inter-item correlation was .22. As disordered eating represents a broad construct, this was an acceptable mean inter-item correlation. For the individual inter-item correlations, values between .15 and .50 are recommended (Briggs & Cheek, 1986; Clark & Watson, 1995). As shown in Table 25, 10 of the 20 items had at least one correlation below this range. In particular, items 11, 14, and 19 had at least one inter-item correlation below .15, item 13 had two, items 16 and 17 had three, items 9 and 10 had four, while item 15 had five. Items 18 and 20 did not correlate well with some of the other items (8 and 7 inter-item correlations below .15, respectively). Only one inter-item correlation (between items 1 and 4) was observed to be greater than .50.

Table 25

Inter-Item Correlations for the Developed Instrument (N = 444)

Item	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1	–																			
2	.28	–																		
3	.45	.34	–																	
4	.54	.41	.42	–																
5	.26	.39	.32	.32	–															
6	.41	.23	.30	.48	.20	–														
7	.43	.26	.27	.46	.18	.34	–													
8	.18	.31	.16	.24	.19	.15	.29	–												
9	.19	.03	.14	.20	.08	.27	.28	.11	–											
10	.17	.14	.16	.21	.15	.13	.08	.11	.19	–										
11	.25	.34	.28	.32	.31	.24	.24	.23	.02	.27	–									
12	.29	.16	.23	.30	.19	.19	.21	.17	.15	.31	.41	–								
13	.49	.26	.26	.48	.21	.33	.36	.11	.11	.21	.28	.38	–							
14	.34	.29	.29	.36	.21	.21	.23	.16	.10	.20	.25	.28	.54	–						
15	.19	.25	.09	.21	.13	.16	.17	.09	-.07	.09	.16	.15	.16	.22	–					
16	.27	.15	.15	.26	.14	.15	.43	.19	.17	.08	.17	.20	.29	.23	-.05	–				
17	.24	.44	.29	.31	.41	.16	.22	.25	.06	.13	.41	.28	.18	.25	.26	.06	–			
18	.22	.14	.08	.15	.14	.04	.15	.12	.15	.15	.17	.15	.09	.09	.21	.14	.22	–		
19	.28	.25	.24	.39	.21	.24	.43	.26	.16	.19	.27	.24	.38	.33	.04	.39	.21	.15	–	
20	.19	.10	.17	.21	.14	.09	.17	.09	.01	.09	.17	.19	.21	.20	.24	.12	.20	.16	.23	–

Internal consistency. Reliability analyses were performed to assess the internal consistency of the developed instrument. Given the dichotomous nature of the instrument, K-R 20 was used. Results revealed a K-R 20 index of .85, suggesting the new instrument had a high level of internal consistency when all 20-items were included (DeVillis, 2012; Kline, 2005). To assess the 'contribution' or 'fit' of each item in the scale, item-total statistics were inspected, indicating that removal of item 9 would marginally increase the K-R 20 index from .851 to .852. The corrected item-total correlations revealed low correlations (< .30) for items 9, 15, 18, and 20, suggesting these four items may not be measuring the same underlying construct.

Principal Components Analysis (PCA). PCA was performed to identify the underlying latent structure of the developed instrument (i.e., whether the scale was uni- or multi- dimensional) and reduce the number of original items. PCA was selected over factor analysis as it has been suggested that PCA potentially provides a clearer understanding of latent structure (Loewenthal, 1996). The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than .20. The overall Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .87, a classification of 'meritorious' according to Kaiser (1974). KMO values for the individual items ranged from .70 to .93, classifications of 'middling' to 'marvelous' (Kaiser, 1974). The overall and individual KMO values suggested the data was suitable for PCA. Bartlett's Test of Sphericity value was statistically significant, indicating the data was likely capable of factorisation. No small communalities (< .30) were revealed.

PCA revealed six components that had eigenvalues greater than one (Kaiser, 1974), which explained 27.5%, 8.0%, 6.1%, 6.0%, 5.7%, and 5.4% of the total variance, respectively. In total, these components accounted for around 58.6% of the variance in the developed instrument. However, only two components contributed at least 8% of the total

variance (Loewenthal, 1999). Only the first component had an eigenvalue greater than two (Linacre, 2014). Visual inspection of the scree plot suggested retention of two components (Cattell, 1966); however, a three-component solution met the interpretability criterion. As such, forced factor extractions for one and two components were performed.

The one-component solution explained 27.5% of the total variance, while the two-component solution explained 35.5%. Varimax orthogonal rotations were employed in both PCAs to aid interpretability. As shown in Table 26, 15 of the 20 items in the one-component PCA exhibited item loadings greater than .40. Items 9, 10, 15, 18, and 20 did not load strongly onto the single factor solution. Notably, these items were also suggested for removal based on the inter-item correlations and item-total statistics.

Table 26

Rotated Structure Matrix for PCA with Varimax Rotation of a One-Component Scale

Developed Instrument Item	Component 1	Communalities
4. Distressed after eating	**.74**	.55
1. Distressed about changes to body/eating	**.68**	.46
13. Disgust with pregnancy body	**.65**	.42
7. Fear of fatness	**.62**	.38
19. Comparison of weight/shape/size/eating	**.59**	.34
14. Overvaluation of weight/shape/size	**.58**	.34
2. Desire to stop bodily changes	**.58**	.33
3. Anxious about eating	**.57**	.33
11. Researching pregnancy weight loss	**.57**	.32
17. Using pregnancy as excuse to avoid food	**.54**	.29
6. Perceived loss of control over eating	**.53**	.29
12. Researching rapid postpartum weight loss	**.53**	.28
5. Rules and conditions connected to eating	**.51**	.26
16. Desire for small pregnancy body	**.44**	.19
8. Repeated/frequent weighing	**.41**	.17
9. Distressed regarding perceived overeating	.39	.16

Note. Factor loadings $\geq$ .40 are in boldface.

Table 26 (continued)

Rotated Structure Matrix for PCA with Varimax Rotation of a One-Component Scale

Developed Instrument Item	Component 1	Communalities
10. Self-induced vomiting	.36	.13
20. Emotional detachment	.36	.13
15. Resent toward baby	.33	.11
18. Desire for baby to be small/petite	.31	.10

The rotated solution in the two-component forced extraction revealed a 'simple structure' (Thurstone, 1947), with 16 of the 20 items exhibiting item loadings greater than .40. As shown in Table 27, items 8, 10, 18, and 20 did not load strongly on the two-component solution. As noted in the previous paragraph, items 10, 18, and 20 had already been indicated for potential removal in the inter-item correlations and item-total statistics. Interpretation of the two-component solution revealed the first component likely reflects cognitions and emotions connected to shape, weight, and eating concerns during pregnancy, while the second component reflects gestational behaviours that result from shape, weight, and eating concerns during pregnancy.

Table 27

Rotated Structure Matrix for PCA with Varimax Rotation of a Two-Component Scale

	Component 1	Component 2	Communalities
7. Fear of fatness	**.69**	.18	.51
16. Desire for small pregnancy body	**.64**	-.02	.41
13. Disgust with pregnancy body	**.61**	.31	.47
19. Comparison of weight/shape/size/eating	**.61**	.22	.42
1. Distressed about changes to body/eating	**.61**	.35	.49
4. Distressed after eating	**.61**	.45	.56
6. Perceived loss of control over eating	**.54**	.22	.33
9. Distressed regarding perceived overeating	**.52**	-.11	.28
14. Overvaluation of weight/shape/size	**.43**	.39	.34

Note. Factor loadings $\geq$ .40 are in boldface.

Table 27 (continued)

Rotated Structure Matrix for PCA with Varimax Rotation of a Two-Component Scale

	Component 1	Component 2	Communalities
17. Using pregnancy as excuse to avoid food	.02	**.74**	.55
2. Desire to stop bodily changes	.16	**.66**	.46
11. Researching pregnancy weight loss	.19	**.62**	.42
5. Rules and conditions connected to eating	.13	**.59**	.37
15. Resent toward baby	-.07	**.54**	.30
3. Anxious about eating	.37	**.44**	.33
12. Researching rapid postpartum weight loss	.34	**.42**	.29
20. Emotional detachment	.15	.36	.15
8. Repeated/frequent weighing	.25	.33	.17
18. Desire for baby to be small/petite	.11	.33	.12
10. Self-induced vomiting	.22	.30	.13

Note. Factor loadings $\geq$ .40 are in boldface.

Item difficulty / variation. To identify which items had the greatest level of item differentiation, the mean score of each item was inspected. The mean of a dichotomous item is equal to the proportion of individuals who endorse the item, thereby revealing the variability of an item (Gregory, 2015; Kline, 2005). Classical test theory suggests .50 represents the optimal level of item difficulty (Furr & Bacharach, 2008; Kline, 2005). Seven items displayed maximum variability (items 1, 3, 6, 7, 9, 16, and 19). Three items displayed low variability (items 10, 15, and 17). See Table 28 for an overview.

Table 28

Means and Standard Deviations for the Developed Instrument Items to Determine Item Difficulty (N = 444)

	Mean (p_i)	*SD*
1. Distressed about changes to body/eating	.44	.50
2. Desire to stop bodily changes	.14	.34
3. Anxious about eating	.37	.48
4. Distressed after eating	.33	.47
5. Rules and conditions connected to eating	.10	.30
6. Perceived loss of control over eating	.39	.49
7. Fear of fatness	.61	.49
8. Repeated/frequent weighing	.22	.41
9. Distressed regarding perceived overeating	.57	.50
10. Self-induced vomiting	.04	.20
11. Researching pregnancy weight loss	.12	.33
12. Researching rapid postpartum weight loss	.18	.38
13. Disgust with pregnancy body	.23	.42
14. Overvaluation of weight/shape/size	.14	.35
15. Resent toward baby	.09	.28
16. Desire for small pregnancy body	.62	.49
17. Using pregnancy as excuse to avoid food	.09	.29
18. Desire for baby to be small/petite	.26	.44
19. Comparison of weight/shape/size/eating	.45	.50
20. Emotional detachment	.16	.37

Item discrimination / effect. To determine how well each item of the developed instrument could discriminate between women with high and low total scores, discrimination indices for each item were calculated using the extreme group method (Furr & Bacharach, 2008; Kline, 2005). That is, participants with the highest and lowest total developed instrument scores were categorised into upper and lower groups. A criterion of 27% was used when creating the extreme groups, as outlined by Cureton (1957) and Kline (2005). After mean scores for each item were calculated for the upper and lower groups, respectively, the

mean score for the lower group was subtracted from the upper group for each item of the developed instrument. A higher discrimination index is reflective of an item discriminating well among individuals with different amounts of the underlying concept of interest (Anastasi & Urbina, 1997). As shown in Table 29, items 10 and 15 had poor discrimination indices, indicating these two items did not discriminate well between those who scored high and low on the developed instrument.

Table 29

Item Discrimination Indices for Each Item Using the Extreme Group Method

Developed Instrument Item		*n*	*M*	*SD*	*D*
1. Distressed about changes to body/eating	Lower group	120	.06	.24	.86
	Upper group	120	.92	.28	
2. Desire to stop bodily changes	Lower group	120	.01	.09	.38
	Upper group	120	.39	.49	
3. Anxious about eating	Lower group	120	.10	.30	.67
	Upper group	120	.77	.43	
4. Distressed after eating	Lower group	120	.01	.09	.84
	Upper group	120	.85	.36	
5. Rules and conditions connected to eating	Lower group	120	.00	.00	.28
	Upper group	120	.28	.45	
6. Perceived loss of control over eating	Lower group	120	.05	.22	.70
	Upper group	120	.75	.44	
7. Fear of fatness	Lower group	120	.08	.26	.88
	Upper group	120	.96	.20	
8. Repeated/frequent weighing	Lower group	120	.04	.20	.39
	Upper group	120	.43	.50	
9. Distressed regarding perceived overeating	Lower group	120	.22	.41	.56
	Upper group	120	.78	.42	
10. Self-induced vomiting	Lower group	120	.00	.00	**.14**
	Upper group	120	.14	.35	
11. Researching pregnancy weight loss	Lower group	120	.00	.00	.33
	Upper group	120	.33	.47	
12. Researching rapid postpartum weight loss	Lower group	120	.02	.13	.45
	Upper group	120	.47	.50	
13. Disgust with pregnancy body	Lower group	120	.00	.00	.62
	Upper group	120	.62	.49	
14. Overvaluation of weight/shape/size	Lower group	120	.01	.09	.42
	Upper group	120	.43	.50	

Note. Boldface is indicative of poor group discrimination. *n* = number in each group, *M* = mean, *SD* = standard deviation, *D* = discrimination index.

Table 29 (continued)

Item Discrimination Indices for Each Item Using the Extreme Group Method

Developed Instrument Item		*n*	*M*	*SD*	*D*
15. Resent toward baby	Lower group	120	.02	.40	**.18**
	Upper group	120	.20	.31	
16. Desire for small pregnancy body	Lower group	120	.23	.42	.67
	Upper group	120	.90	.30	
17. Using pregnancy as excuse to avoid food	Lower group	120	.00	.00	.28
	Upper group	120	.28	.45	
18. Desire for baby to be small/petite	Lower group	120	.07	.25	.36
	Upper group	120	.43	.50	
19. Comparison of weight/shape/size/eating	Lower group	120	.08	.28	.77
	Upper group	120	.85	.36	
20. Emotional detachment	Lower group	120	.04	.20	.30
	Upper group	120	.34	.48	

Note. Bolding is indicative of poor group discrimination. *n* = number in each group, *M* = mean, *SD* = standard deviation, *D* = discrimination index.

Item performance. A series of chi-square tests of association were performed with the developed instrument items and the three EDE-Q clinical cut-offs. The first series of chi-square analyses examined the developed instrument items and the EDE-Q with a cut-off of 2.80. The second series of chi-square analyses examined the developed instrument items and the EDE-Q with a cut-off of 2.50. The third series of chi-square analyses examined the developed instrument items and the EDE-Q with a cut-off of 2.30. When a participant equaled or exceeded the respective EDE-Q cut-off, a value of 1 was assigned, representative of case identification. Participants with scores below the cut-off were assigned a value of 0.

Results of the first chi-square analyses revealed statistically significant associations between 19 of the developed instrument items and the EDE-Q clinical cut-off ($\geq$ 2.80). Item 19 of the developed instrument did not display a statistically significant association with the EDE-Q clinical cut-off ($\geq$ 2.80). Inspection of the cross-tabulation tables revealed six items where less than 50 percent of women in the EDE-Q clinical range responded affirmatively, suggesting poor item performance. These items were 8, 10, 15, 17, 18, and 20 (see Table 30 on the following page).

Results of the second chi-square analyses revealed statistically significant associations between 19 of the developed instrument items and the EDE-Q clinical cut-off (≥ 2.50). Item 19 of the developed instrument again did not display a statistically significant association with the EDE-Q clinical cut-off (≥ 2.50). Inspection of the cross-tabulation tables revealed five items where less than 50 percent of women in the EDE-Q clinical range responded affirmatively, suggesting poor item performance. These items were 8, 10, 15, 17, and 20 (see Table 30).

Results of the final chi-square analyses revealed statistically significant associations between the 20 items of the developed instrument and the EDE-Q clinical cut-off (≥ 2.30). Inspection of the cross-tabulation tables revealed six items where less than 50 percent of women in the EDE-Q clinical range responded affirmatively, suggesting poor item performance. These items were 5, 8, 10, 15, 17, and 20 (see Table 30).

Table 30

Item Performance Analyses for Each Item at Three EDE-Q Cut-Off Thresholds

Item	Endorsement from women in clinical range (≥ 2.8) (*n* = 35)			Endorsement from women in clinical range (≥ 2.5) (*n* = 38)			Endorsement from women in clinical range (≥ 2.3) (*n* = 45)		
	f	%	φ	*f*	%	φ	*f*	%	φ
1	32	91.4%	.28	35	92.1%	.30	42	93.3%	.34
2	18	51.4%	.32	20	52.6%	.35	23	51.1%	.37
3	30	85.7%	.30	33	86.8%	.32	37	82.2%	.32
4	34	97.1%	.40	37	97.4%	.42	43	95.6%	.45
5	18	51.4%	.40	19	50.0%	.40	21	**46.7%**	.41
6	32	91.4%	.31	35	92.1%	.33	40	88.9%	.34
7	34	97.1%	.22	37	97.4%	.23	44	97.8%	.25
8	14	**40.0%**	.13	15	**39.5%**	.13	20	**44.4%**	.19
9	25	71.4%	.09	27	71.1%	.09	33	73.3%	.11
10	9	**25.7%**	.31	10	**26.3%**	.33	11	**24.4%**	.34

Note. *f* = number of women in the clinical range who endorsed the item. Φ = Phi and Cramer's V (effect size). Bolding denotes items where less than 50% of women in the clinical responded affirmatively.

Table 30 (continued)

Item Performance Analyses for Each Item at Three EDE-Q Cut-Off Thresholds

Item	Endorsement from women in clinical range (≥ 2.8) (*n* = 35)			Endorsement from women in clinical range (≥ 2.5) (*n* = 38)			Endorsement from women in clinical range (≥ 2.3) (*n* = 45)		
	f	%	φ	*f*	%	φ	*f*	%	φ
11	21	60.0%	.43	23	60.5%	.46	24	53.3%	.43
12	22	62.9%	.35	24	63.2%	.36	28	62.2%	.39
13	29	82.9%	.42	32	84.2%	.45	38	84.4%	.50
14	24	68.6%	.46	26	68.4%	.48	29	64.4%	.49
15	7	**20.0%**	.12	9	**23.7%**	.16	10	**22.2%**	.16
16	30	85.7%	.14	33	86.8%	.16	39	86.7%	.17
17	12	**34.3%**	.25	14	**36.8%**	.29	16	**35.6%**	.31
18	16	**45.7%**	.13	19	50.0%	.16	23	51.1%	.19
19	32	91.4%	.27	35	92.1%	.29	42	93.3%	.33
20	13	**18.6%**	.17	14	**36.8%**	.18	18	**40.0%**	.22

Note. *f* = number of women in the clinical range who endorsed the item. Φ = Phi and Cramer's V (effect size). Bolding denotes items where less than 50% of women in the clinical responded affirmatively.

Items removed. A summary of statistical characteristics for each item of the developed instrument is shown in Table 31. Overall, six items were removed, resulting in a 14 item self-report scale. Given most of the items were related to cognitive or affective symptomatology of disordered eating, the instrument was labeled the "Disordered Eating Attitudes in Pregnancy Scale" (DEAPS). Scoring for the DEAPS involved summing the number of affirmative responses (*yes* = 1, *no* = 0), with total scores ranging from zero to 14. A copy of the 14 item DEAPS can be found in Appendix F. Part II of this study aimed to determine a suitable cut-off score on the DEAPS for potential case identification.

Table 31

Summary of Item Reduction Analyses for the DEAPS, Including the Decision to Retain or Delete Individual Items

DEAPS	**Inter-item correlations below .15**	**Item-total correlation below .30**	**K-R 20 would improve**	**Does not fit one component solution**	**Does not fit two component solution**	**Item difficulty below .25**	**Low item discrimination Index**	**Poor item performance with EDE-Q**	**Decision**	**Rationale (to retain or remove, if necessary)**
Item 1									Retain	
Item 2						X			Retain	Pertinent to the construct of disordered eating
Item 3									Retain	
Item 4									Retain	
Item 5						X		X	Retain	Pertinent to the construct of disordered eating
Item 6									Retain	
Item 7									Retain	
Item 8					X			X	Delete	Poor performance with EDE-Q
Item 9	X	X	X	X					Retain	Pertinent to the construct of disordered eating
Item 10	X			X	X	X	X	X	Delete	Poor performance across various item reduction tests
Item 11	X					X			Retain	Fair performance with EDE-Q
Item 12						X			Retain	Fair performance with EDE-Q
Item 13	X								Retain	Good performance with EDE-Q
Item 14	X					X			Retain	Fair performance with EDE-Q
Item 15	X	X		X		X	X	X	Delete	Poor performance across various item reduction tests
Item 16	X								Retain	Good performance with EDE-Q
Item 17	X					X		X	Delete	Poor performance with EDE-Q
Item 18	X	X		X	X			X	Delete	Poor performance with EDE-Q
Item 19	X								Retain	
Item 20	X	X		X	X	X		X	Delete	Poor performance across various item reduction tests

Part II – Psychometric Analyses

A description of each psychometric property can be found in Chapter 4 (pp. 148-151).

Internal consistency. To explore the reliability of the final DEAPS instrument (i.e., 14 items), in comparison to the EDE-Q and SCOFF, internal consistency estimates for each measure were calculated. A conceptual description of internal consistency can be found in Chapter 4. As shown in Table 32, the DEAPS displayed a high level of internal consistency across the two subsamples, similar to the EDE-Q, while poor internal consistency was revealed for the SCOFF.

Table 32

Internal Consistency Analyses for the DEAPS, EDE-Q, and SCOFF

	Number of items	Subsample A (n = 222)	Subsample B (n = 222)	Total Sample (N = 444)
DEAPS[a]	14	.85	.84	.85
EDE-Q[b]	22	.95	.94	.95
SCOFF[a]	5	.43	.50	.47

Note. DEAPS = Disordered Eating Attitudes in Pregnancy Scale. EDE-Q = Eating Disorder Examination Questionnaire. SCOFF = Sick Control One Fat Food Questionnaire. [a] K-R 20 indices. [b] Cronbach alpha.

Content validity. Prior to the psychometric evaluation of the DEAPS, the instrument was reviewed by 15 subject matter experts for construct relevance; item clarity, conciseness, and phrasing; and potential implementation issues. Feedback from the 15 subject matter experts indicated that 17 of the 19 DEAPS items were relevant to the construct of interest, presented in a clear and concise manner, appropriately phrased, and non-threatening in nature. The instrument was then piloted with a sample of 12 pregnant women at various gestational stages. Participants in the piloting study reported the response format was easy to navigate, the level of information requested was reasonable, and almost all the items were phrased clearly and appropriately. See Steps 4 and 5 in Phase I for further information.

In addition to the feedback obtained from Phase I, participants in Phase II were asked to indicate how comfortable they felt answering the questions on each psychometric scale using a 6-point Likert scale (1 = *very uncomfortable* to 6 = *very comfortable*). As shown in Table 33, participants' comfort level in completing the three questionnaires ranged from slightly comfortable (EDE-Q) to quite comfortable (DEAPS and SCOFF). Paired samples *t*-tests revealed that, across the total sample, participants were more comfortable completing the DEAPS $t(443) = 17.79$, $p = < .001$ and SCOFF $t(443) = 19.85$, $p = < .001$, compared to the EDE-Q. No significant differences in comfort/acceptability were revealed between the DEAPS and SCOFF questionnaire. Additionally, no significant differences were revealed for the comfort/acceptability of individual scales across Subsample A and Subsample B.

Table 33

Scale Comfort/Acceptability Ratings for the DEAPS, EDE-Q, and SCOFF

	Subsample A	Subsample B	Total Sample	Subsample Comparison
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*t*-test
DEAPS	5.10 (1.02)	5.17 (.98)	5.14 (1.01)	$t(442) = -.71$, $p = .481$
EDE-Q	4.64 (1.04)	4.48 (1.03)	4.56 (1.04)	$t(442) = 1.65$, $p = .099$
SCOFF	5.31 (.94)	5.25 (1.03)	5.28 (.98)	$t(442) = .58$, $p = .563$

Note. DEAPS = Disordered Eating Attitudes in Pregnancy Scale. EDE-Q = Eating Disorder Examination Questionnaire. SCOFF = Sick Control One Fat Food Questionnaire.

Construct validity. An additional PCA was performed for the refined DEAPS scale, confirming the unidimensional latent structure (see Table 34).

Table 34

Component Matrix for PCA with Varimax Rotation on the One-Component DEAPS

DEAPS item	Subsample A	Subsample B	Total Sample
1. Distressed about changes to body/eating	.73	.68	.71
2. Desire to stop bodily changes	.57	.53	.55
3. Anxious about eating	.61	.57	.59
4. Distressed after eating	.79	.74	.76
5. Rules and conditions connected to eating	.47	.49	.48
6. Perceived loss of control over eating	.54	.60	.57
7. Fear of fatness	.63	.65	.64
9. Distressed regarding perceived overeating	.38	.44	.41
11. Researching pregnancy weight loss	.58	.51	.54
12. Researching rapid postpartum weight loss	.51	.52	.52
13. Disgust with pregnancy body	.72	.66	.69
14. Overvaluation of weight/shape/size	.61	.57	.58
16. Desire for small pregnancy body	.45	.51	.48
19. Comparison of weight/shape/size/eating	.59	.61	.60
KMO value	.86	.85	.88
Bartlett's test	< .001	< .001	< .001

Convergent validity. Pearson product-moment bivariate correlations were performed using the DEAPS total score, the EDE-Q global score, and the SCOFF total score. As shown in Table 35, the DEAPS demonstrated a strong correlation with the EDE-Q, providing support for convergent validity, and therefore construct validity. A moderate correlation between the DEAPS and SCOFF was also revealed, with a similar correlation observed between the EDE-Q and SCOFF.

Table 35

Summary of Inter-Correlations Between the DEAPS, EDE-Q, and SCOFF (*N = 444)*

	DEAPS	EDE-Q	SCOFF
Subsample A (*n* = 222)			
DEAPS	–		
EDE-Q	.82***	–	
SCOFF	.64***	.65***	–
Subsample B (*n* = 222)			
DEAPS	–		
EDE-Q	.78***	–	
SCOFF	.63***	.63**	–
Total Sample (*N* = 444)			
DEAPS	–		
EDE-Q	.80***	–	
SCOFF	.63***	.64***	–

Note. ****p* < .001 (two-tailed)

Criterion-related validity. As shown in Table 36, ROC analyses using the three EDE-Q cut-off points revealed an optimal compromise between sensitivity and specificity on the DEAPS in Subsample A was achieved at a total score of 7.50 (Se = 1.00, Sp = .83, AUC = .97). Using this cut-off score, 49 women (22.1%) obtained a positive screen for possible disordered eating symptomatology on the DEAPS, while 173 (77.9%) obtained a negative screen. Mann-Whitney U tests confirmed that women who obtained a positive screen were more likely to respond affirmatively to each of the DEAPS items.

In Subsample B, ROC analyses using the three EDE-Q cut-off points revealed an optimal compromise between sensitivity and specificity on the DEAPS was also achieved at a total score of 8.00 (Se = .86, Sp = .86, AUC = .94). Using this cut-off, 48 women (21.6%) obtained a positive screen for possible disordered eating symptomatology on the DEAPS, while 174 (78.4%) obtained a negative screen. Mann-Whitney U tests confirmed that women who obtained a positive screen were more likely to respond affirmatively to each of the

DEAPS items. The DEAPS also produced appropriate LRs in both subsamples for a screening instrument (McGee, 2002).

Table 36

Results of the ROC Analyses for the DEAPS Across Subsample A and Subsample B

	Subsample A (n = 222)			Subsample B (n = 222)		
DEAPS (≥ 8)	EDE-Q (≥ 2.3)	EDE-Q (≥ 2.5)	EDE-Q (≥ 2.8)	EDE-Q (≥ 2.3)	EDE-Q (≥ 2.5)	EDE-Q (≥ 2.8)
Sensitivity	1.00	1.00	1.00	.88	.86	.85
Specificity	.80	.84	.84	.86	.86	.85
AUC	.97	.97	.97	.94	.94	.93
95% CI	[.94-.99]	[.95-.99]	[.94-.99]	[.88-.99]	[.88-1.00]	[.86-1.00]
LR+	1.25	1.10	1.20	1.01	1.01	1.00
LR-	0	0	0	.92	.94	.98

Note. EDE-Q = Eating Disorder Examination Questionnaire. DEAPS = Disordered Eating Attitudes in Pregnancy Scale. Se = sensitivity. Sp = specificity. AUC = area under the curve. CI = confidence interval. LR+ = positive likelihood ratio. LR– = negative likelihood ratio.

Sensitivity and specificity values were also calculated for the SCOFF at the recommended cut-off score of two (see Table 37). In Subsample A, overall sensitivity was .81 and overall specificity was .86 (AUC = .90). In Subsample B, overall sensitivity was .73 and overall specificity was .82 (AUC = .87). Mann-Whitney U tests also confirmed that women who obtained a positive screen on the SCOFF were more likely to respond affirmatively to each of the SCOFF items. Unlike the DEAPS, the SCOFF produced suboptimal LRs in both subsamples for a screening instrument (McGee, 2002). The high LR- values indicate a proportion of women with disordered eating incorrectly obtained a negative screening result.

Table 37

Results of the ROC Analyses for the SCOFF Across Subsample A and Subsample B

	Subsample A (n = 222)			Subsample B (n = 222)		
SCOFF (≥ 2)	EDE-Q (≥ 2.3)	EDE-Q (≥ 2.5)	EDE-Q (≥ 2.8)	EDE-Q (≥ 2.3)	EDE-Q (≥ 2.5)	EDE-Q (≥ 2.8)
Sensitivity	.76	.81	.87	.71	.73	.75
Specificity	.87	.85	.86	.82	.82	.81
AUC	.87	.91	.92	.87	.87	.87
95% CI	[.79-.95]	[.85-.97]	[.86-.97]	[.81-.93]	[.81-.94]	[.81-.93]
LR+	.88	.95	1.01	.87	.89	.92
LR-	1.78	1.28	.92	1.60	1.48	1.33

Note. EDE-Q = Eating Disorder Examination Questionnaire. SCOFF = Sick Control One Fat Food Questionnaire. Se = sensitivity. Sp = specificity. AUC = area under the curve. CI = confidence interval. LR+ = positive likelihood ratio. LR– = negative likelihood ratio.

Part III – Prevalence of Disordered Eating in the Sample

Prevalence across the different scales. As shown in Table 38, the prevalence of disordered eating in the total sample ranged from 8 to 10 percent when using the EDE-Q cut-offs. When using the DEAPS and SCOFF cut-offs, the prevalence of disordered eating in the sample increased up to 22 percent. Prevalence rates for each instrument were also consistent across Subsample A and Subsample B, with no significant differences observed.

Table 38

Prevalence of Disordered Eating in the Sample According to the Various Instruments

	Subsample A	Subsample B	Total Sample	Subsample comparison
	n (%)	n (%)	n (%)	χ^2
DEAPS	49 (22.1%)	48 (21.6%)	97 (21.8%)	$\chi^2(1) = .01$, $p = .909$
EDE-Q (≥ 2.3)	21 (9.5%)	24 (10.8%)	45 (10.1%)	$\chi^2(1) = .22$, $p = .637$
EDE-Q (≥ 2.5)	16 (7.2%)	22 (9.9%)	38 (8.6%)	$\chi^2(1) = 1.04$, $p = .309$
EDE-Q (≥ 2.8)	15 (6.8%)	20 (9.0%)	35 (7.9%)	$\chi^2(1) = .76$, $p = .379$
SCOFF	43 (19.4%)	53 (23.9%)	96 (21.6)	$\chi^2(1) = 1.33$, $p = .249$

Note. DEAPS = Disordered Eating Attitudes in Pregnancy Scale. EDE-Q = Eating Disorder Examination Questionnaire. SCOFF = Sick Control One Fat Food Questionnaire. n = number. χ^2 = chi-square statistics. p = significance value.

Endorsement of individual DEAPS items across the total sample. Analysis of the individual DEAPS items for the total sample (see Table 38) revealed that just less than three quarters worried they had or would become "fat" during pregnancy and desired their pregnancy body to be "small". More than half felt distressed and uncomfortably full after eating a large amount of food. Just under half the sample felt distressed about the changes to their body or eating during pregnancy, found themselves comparing their body weight/shape/size or eating habits to other women, felt like they had lost control over their body in the previous month, and felt anxious about eating. Around one third felt distressed after eating because of the effect on their body weight/shape/size. More than a quarter felt disgusted or ashamed with their pregnancy body and invested considerable time researching methods to rapidly lose weight in the postpartum period. A quarter believed their pregnancy weight/shape/size was important when evaluating self-worth. Less than a quarter also reported engaging in behaviours to prevent pregnancy-related bodily changes (e.g., dietary restriction, purging, excessive exercise), investing significant time to research methods to prevent pregnancy-related weight gain, and following dietary rules.

Endorsement of individual DEAPS items in positive screen cases. Analysis of the individual DEAPS items endorsed by women who obtained a positive screening result (i.e., a total DEAPS score of 8 or more) revealed that almost all the women who obtained a positive screening result worried they had or would become fat during pregnancy, felt distressed about the changes to their body and eating patterns, wanted their pregnancy body to be small, and felt distressed after eating due to the potential impact on their weight and shape (see Table 39). More than three quarters found themselves comparing their body weight/shape/size or eating habits to other women (pregnant and non-pregnant), experiencing anxiety in relation to food, felt like they had lost control over their body in the previous month, and felt distressed and uncomfortably full after eating a large amount of food. Almost

one third felt disgusted or ashamed of their pregnancy body. Almost half had invested a significant amount of time researching methods to rapidly lose weight in the postpartum period, reported engaging in behaviours to prevent pregnancy-related bodily changes (e.g., dietary restriction, purging, excessive exercise), and believed their pregnancy weight/shape/size was important when evaluating their self-worth as a mother. More than one third indicated they had spent considerable time researching ways to prevent pregnancy-related weight gain and following strict dietary compensation rules.

Table 39

Frequency Analysis of Positive ('Yes') Responses for each DEAPS Item for all Participants, Compared to Those Who Obtained a Positive Screen

DEAPS	Subsample A		Subsample B		Total sample	
Over the past month…	All	Positive	All	Positive	All	Positive
1. Distressed about changes to body/eating	42.3%	93.9%	45.0%	97.9%	43.7%	95.9%
2. Desire to stop pregnancy-related changes	14.0%	44.9%	13.1%	47.9%	13.5%	46.4%
3. Anxious about eating	35.1%	85.7%	38.3%	85.4%	36.7%	85.6%
4. Distressed after eating	29.3%	89.8%	36.0%	93.8%	32.7%	91.8%
5. Rules and conditions connected to eating	9.0%	30.6%	36.0%	35.4%	10.1%	33.0%
6. Perceived loss of control over eating	38.3%	81.6%	40.5%	83.3%	39.4%	82.5%
7. Fear of fatness	59.9%	98.0%	62.6%	97.9%	61.3%	97.9%
8. Distressed regarding perceived overeating	56.8%	75.5%	56.3%	81.3%	56.5%	78.4%
9. Researching pregnancy weight loss	8.6%	32.7%	15.3%	41.7%	11.9%	37.1%
10. Researching rapid postpartum weight loss	16.2%	46.9%	19.4%	52.1%	17.8%	49.5%
11. Disgust with pregnancy body	23.0%	71.4%	22.1%	62.5%	22.5%	67.0%
12. Overvaluation of weight/shape/size	13.1%	46.9%	15.3%	50.0%	14.2%	48.5%
13. Desire for small pregnancy body	57.7%	89.8%	66.2%	93.8%	61.9%	91.8%
14. Comparison of weight/shape/size/eating	42.8%	91.8%	47.3%	85.4%	45.0%	88.7%

Note. All = total sample (N = 444). Positive = women who obtained a total DEAPS score of $\geq$ 8 (n = 97).

Group differences. To explore demographic differences between those who obtained a positive screen on the DEAPS versus those who obtained a negative screen, a series of independent samples t-tests (continuous demographic variables) and chi-square analyses (nominal demographic variables) were performed for each subsample. To control for family wise error, resultant from multiple comparisons, a Šidák-Bonferroni correction was applied. This reduced the alpha level from .05 to .004.

The independent samples t-tests (see Table 40) revealed no significant differences between the positive and negative screening groups in number of children, total pregnancies, and gestational week. Significant differences were observed for age and BMI; however, this did not remain significant once the Šidák-Bonferroni correction was applied.

Table 40

Demographic Overview and Comparison Based on the DEAPS Screening Outcome Groups

Variable	Non-cases (*n* = 347)	Cases (*n* = 97)		
	M (*SD*)	*M* (*SD*)	*t*	*p*
Age	29.59 (4.69)	28.13 (5.54)	2.59	.010*
No. of children	.53 (1.07)	.44 (.90)	.76	.450
Total pregnancies	1.84 (1.55)	1.90 (2.13)	-.27	.787
Gestational week	23.32 (9.74)	21.71 (9.12)	1.46	.146
Pre-preg BMI	25.57 (5.66)	27.01 (6.74)	-2.21	.028*
Pregnancy BMI	27.63 (6.74)	29.18 (6.79)	-2.22	.028*

Note. *t* = t-value from independent samples t-test. *p* = significance value from the independent samples t-test conducted.
* = significance value did not remain significant once Šidák -Bonferroni correction was applied.

A series of chi-square analyses revealed no significant differences between the positive and negative screening groups in terms of ethnicity $\chi^2(8) = 6.69$, $p = .570$, education $\chi^2(9) = 16.58$, $p = .056$, employment status $\chi^2(7) = 10.96$, $p = .146$, conception duration $\chi^2(4) = 1.71$, $p = .789$, nature of conception (e.g., unassisted vs. assisted) $\chi^2(1) = 2.13$, $p = .145$, or pregnancy trimester $\chi^2(2) = 3.66$, $p = .160$. Significant differences were observed between the

positive and negative screening groups in relation to relationship status $\chi^2(5) = 12.10$, $p = .033$, conception context (e.g., planned vs. unplanned) $\chi^2(3) = 13.18$, $p = .004$, and self-reported history of disordered eating $\chi^2(1) = 42.10$, $p = < .001$. The difference in relationship status did not remain significant once the Šidák-Bonferroni correction was applied. A greater number of women who obtained a positive screen on the DEAPS reported the index pregnancy was unplanned compared to those who obtained a negative screen. A greater number of women who obtained a positive screen on the DEAPS also reported a history of disordered eating compared to those who obtained a negative screen.

Additionally, of those who screened positive for disordered eating using the DEAPS, over half reported no history of disordered eating symptomatology, three quarters were receiving public maternity care, more than half were nulliparous (i.e., pregnant for the first time), and half were in their second trimester (see Table 41).

Table 41

Obstetric History of Women Who Obtained a Positive Screening Outcome Using the DEAPS

Variable	*n* (%)
Hx of ED or disordered eating	
Yes	41 (42.3%)
No	55 (57.3%)
Maternity care	
Public	74 (76.3%)
Private	23 (23.7%)
First pregnancy	60 (61.9%)
Trimester	
1st	19 (19.6%)
2nd	50 (51.5%)
3rd	28 (28.9%)
Total DEAPS score	9.94 ± 1.78 (range = 8-14)

Note. Hx = history. Figures for the total DEAPS score represents the mean and standard deviation (±)

Discussion

The aim of the current study was to develop a pregnancy-specific disordered eating instrument that could be easily implemented, administered, scored, and interpreted by health professionals involved in antenatal care. Following comprehensive development and a two-stage refinement process, the final instrument consisted of 14 self-report items scored in a dichotomous format (i.e., yes/no). As the instrument was more reflective of cognitive and affective symptomatology of disordered eating, conceptualised as subclinical levels of ED symptoms, the instrument was labeled the Disordered Eating Attitudes in Pregnancy Scale (DEAPS). Overall, the DEAPS was found to be a psychometrically sound instrument when compared to the EDE-Q and the SCOFF, with a high level of acceptability among pregnant women.

Psychometric Properties of the DEAPS, EDE-Q, and SCOFF

Consistent with the first hypothesis of the study, the DEAPS demonstrated a high level of internal consistency across subsamples A and B, similar to the EDE-Q. The internal consistency of the DEAPS was also considerably greater than the internal consistency of the SCOFF. The low internal consistency of the SCOFF is, however, similar with previous internal consistency estimates for the instrument in non-pregnant populations (Campo-Arias et al., 2006; Garcia et al., 2011; Hansson et al., 2015; Leung et al., 2015; Mond et al., 2008; Pannocchia et al., 2011) and likely explained by the small number of dichotomous items. With the relationship between scale length and reliability in mind, the high internal consistency of the DEAPS, as a screening instrument with only 14 items, is even more promising.

In terms of validity, the DEAPS demonstrated a high positive correlation with the EDE-Q and a moderate positive correlation with the SCOFF across subsamples A and B, consistent with expectations (hypothesis two). Such correlations provide evidence of

convergent validity, a subset of construct validity. The DEAPS also produced stronger convergent validity with the EDE-Q, compared to the convergent validity between the SCOFF and EDE-Q or the DEAPS and SCOFF. The lower, yet still moderate, correlation of both instruments (EDE-Q and DEAPS) with the SCOFF perhaps suggests slightly different elements of disordered eating symptomatology are assessed by the SCOFF. For example, some of the SCOFF items focus more on behavioural ED symptoms, whereas the DEAPS items appear to have a stronger focus on cognitive and affective ED symptoms, similar to the EDE-Q which has three (of four) subscales assessing cognitive and affective ED symptomatology. Content validity of the DEAPS was also high based on feedback from subject matter experts and pre-piloting with a small subsample of pregnant women. This is not unexpected given expert-derived evidence from Studies 1 and 2 were used to construct the DEAPS items.

Consistent with hypothesis three, the DEAPS displayed a strong level of criterion-related validity against the EDE-Q across subsamples A and B, with an overall sensitivity of 93.2 percent, specificity of 84.2 percent, and an AUC value of .95. These results indicate the DEAPS had excellent accuracy in correctly identifying cases and non-cases of disordered eating in the sample (Fischer, Bachmann, & Jaeschke, 2003). Psychometric literature suggests that a screening test should have at least 80 percent sensitivity and specificity to justify its use (Gregory, 2015). The overall AUC value produced by the DEAPS is also superior to most screening instruments in psychology (Hempel, Buck, Cima, & van Marle, 2013; Jokinen, Nordström, & Nordström, 2008; Kills Small, Simons, & Stricherz, 2007; Mond et al., 2008; Parker et al., 2008) and similar to depression screening tools used in antenatal care (Bunevicius, Kusminskas, Pop, Pedersen, & Bunevicius, 2008; Choi et al., 2012; Chorwe-Sungain & Chipps, 2017). The consistency in reliability and validity data across the primary development and cross-validation samples (subsample A and B,

respectively) also suggests the strong psychometric properties of the DEAPS were not distorted by chance effects, providing preliminary evidence for the stability and robustness of the DEAPS (DeVellis, 2012; Gregory, 2015). While there is no literature outlining an acceptable level of validity shrinkage, the general rule is that less shrinkage is ideal (Miller, Lovler, & McIntire, 2012), as demonstrated in the present study.

In comparison to the DEAPS, the SCOFF had an overall sensitivity of 77.2 percent and specificity of 83.8 percent. While these sensitivity and specificity estimates are somewhat consistent with the performance of the SCOFF in a non-pregnant context (see Botella et al., 2013, for a review), the DEAPS performed better than the SCOFF when identifying symptoms of disordered eating in the sample. Notably, the SCOFF did not identify 22.8 percent of women with disordered eating symptoms, compared to 6.8 percent in the DEAPS. This is a considerable difference in proportion of false negatives, especially when the specificity of the DEAPS was also high. As noted in Chapter 4, issues with content validity may explain the lower sensitivity of the SCOFF. Specifically, the generic and broad item content intended for non-pregnant samples may have resulted in women attributing concerning symptoms to a normative pregnancy experience. For example, the Professional panel in Chapter 3 noted the item "Do you worry you have lost control over how much you eat?" could be ascribed to pregnancy-related appetite increases due to hormonal fluctuations or maternal and/or foetal nutritional needs (King, 2000; Patil, 2012). Additionally, the SCOFF does not contain items relating to unique symptoms of disordered eating in pregnancy.

While the specificity of the SCOFF was promising, the suboptimal sensitivity estimate of the SCOFF is concerning as symptoms of disordered are not typically disclosed openly by women in obstetric care (Franko & Spurrell, 2000; Franko & Walton, 1993; Freizinger et al., 2010; Hollifield & Hobdy, 1990; Morgan, 1997; Newton & Chizawsky,

2006; Tierney et al., 2013). Additionally, such symptoms may not be easily detected without robust psychometric instruments (Easter, 2015; Leddy et al., 2009). As the SCOFF estimated a similar prevalence of disordered eating in the sample as the DEAPS, this reflects the SCOFF is producing a higher number of false positives, as confirmed by the suboptimal negative likelihood ratios. Timeframe differences for each measure may explain the discrepancy in accuracy. For example, the DEAPS specifically asked women to consider each item 'over the past month', as per the consensus recommendation of the Delphi studies in Chapter 1. A similar reference period was used by the EDE-Q (i.e., the previous 28 days). The SCOFF, however, has no specific timeframe, therefore it is unknown what time period women used when rating each SCOFF item. For example, it is possible women might have responded 'yes' to an item if they had experienced the symptom previously; however, this experience may not have been during pregnancy, resulting in a false positive. The brevity of the SCOFF has always been one its main advantages, with administration taking less than 30 seconds; however, at 14 items and less than two minutes to administer, the DEAPS has a similar level of brevity and ease of administration, with superior accuracy.

Partially consistent with hypothesis four, participants reported a similar level of acceptability and comfort when completing the SCOFF and the DEAPS. Inconsistent with this hypothesis, the EDE-Q was perceived to be less comfortable to complete than the SCOFF and the DEAPS. This difference could be attributed to the length of the EDE-Q, in concert with the question content, recall required, and the various Likert rating scales utilised throughout the EDE-Q. Combined with the results of the consumer Delphi study in Chapter 3, which revealed consumers did not perceive longer self-report questionnaires to be practical, this finding suggests the EDE-Q may not be a suitable instrument for screening purposes; however, it may be useful as a follow-up assessment tool. Additional research to

explore the psychometric properties of the EDE-Q is required to explore the performance and utility of the EDE-Q in pregnancy.

Prevalence of Disordered Eating in the Sample

Depending on the psychometric instrument used, the prevalence of disordered eating symptoms in the current sample ranged from 8 to 10 percent when using the EDE-Q cut-offs to 22 percent when using the DEAPS and SCOFF cutoffs. This suggests a considerable proportion of women are experiencing clinically significant disordered eating symptoms during pregnancy. In comparison to previous research, the prevalence of disordered eating in Subsample A (22.1%) and Subsample B (21.6%), as estimated by the DEAPS, was similar to Broussard (2012) who reported a prevalence of 27.8 percent. The DEAPS estimate was, however, higher than prevalence rates reported in other studies, which have ranged from less than one percent to 8.5 percent (Easter et al., 2013; Fairburn, Stein, & Jones, 1992; Kelly et al., 2001; Micali et al., 2007; Mohamadirizi et al., 2015; Pettersson et al., 2016; Soares et al., 2009; Turton, et al., 1999). While methodological differences such as the psychometric instrument administered to identify/measure such symptoms contribute greatly to this discrepancy, increased sensitivity to identify concerns due to the pregnancy-specific nature of the DEAPS may also explain such differences; however, additional research using the DEAPS is required to confirm this.

Using the EDE-Q ($\geq$ 2.80 cutoff), the current study also displayed a much greater prevalence of disordered eating in Subsample A (6.8%) and Subsample B (9.0%) compared to Pettersson et al. (2016) which revealed a prevalence of 3.3 percent. The nature of questionnaire administration is a likely explanation for this difference, with the current study employing anonymous online EDE-Q completion, while Pettersson et al. (2016) used face-to-face administration by a midwife. Given the stigma associated with disordered eating, which is likely to be exacerbated in the context of pregnancy, participants in Pettersson et al. (2016)

may have been reluctant to disclose certain symptoms, resulting in a lower prevalence rate. The negative impact of stigma on symptom disclosure is well documented in perinatal mental health (Highet et al., 2014), particularly in the field of disordered eating (Franko & Spurrell, 2000; Franko & Walton, 1993; Freizinger et al., 2010; Hollifield & Hobdy, 1990; Morgan, 1997; Newton & Chizawsky, 2006; Tierney et al., 2013).

Findings of the current study also revealed no significant differences between women who obtained a positive screen on the DEAPS versus those who obtained a negative screen on most sociodemographic variables, obstetric historical factors, stage of pregnancy, or BMI. This is important as it suggests as absence of bias according to these characteristics. The circumstances/context surrounding conception was the only significant difference observed, with women who obtained a positive screen on the DEAPS more frequently reporting the index pregnancy was either unplanned (but not unexpected) or completely unintended than women who obtained a negative screen. A similar finding was also revealed in the MoBa cohort study, whereby mothers with an ED, particularly AN, were significantly more likely to report the index pregnancy was unintended compared to mothers without an ED (Bulik et al., 2010). It has been suggested the two-fold increased risk of unintended pregnancies in women with EDs, particularly low weight spectrum conditions, may be due to women (and potentially health professionals) incorrectly assuming conception cannot occur in the absence of menstruation (or irregular menstruation), leading to lower or less stringent contraception use (Bulik et al., 2010). For women with pre-existing disordered eating symptoms prior to pregnancy, similar misconceptions may have partially contributed to the greater proportion of unplanned pregnancies in the positive screening group. It is also possible that for some women, the unintended nature of the pregnancy may have contributed to the onset of disordered eating symptomatology to cope with the unexpected social, occupational, relational, and financial changes. Regardless of the variable relationship direction, unintended

pregnancies have been associated with a number of maternal and infant risks (Bahk et al., 2015; Gipson et al., 2008; Klima, 1998; Sawyer et al., 2001), before the potential impact of disordered eating on the pregnancy process is considered. Additional research in this area is required to improve scientific understanding of this relationship and possible prevention strategies.

Providing further support for the validity of the DEAPS, women with a history of disordered eating were significantly more likely to obtain a positive screen on the DEAPS. This is not an unexpected finding and is consistent with observations in previous literature (Blais et al., 2000; Crow et al., 2008; Micali, 2010; Micali et al., 2007; Tierney et al., 2013). Some women may conceive with active disordered eating symptomatology; however, pregnancy is also a period of remarkable changes to a woman's body, eating habits, emotions, and social identity (Franko & Walton, 1993), which may cause previous disordered eating symptoms to resurface. It is well known that the perinatal period, inclusive of pregnancy and postpartum, is a time when dormant psychological issues often re-emerge (Franko & Walton, 1993).

As noted in Chapter 1, pregnancy is also a period of risk for women previously unaffected by disordered eating, particularly for the development of binge eating behaviours (Bulik et al., 2007; Fairburn et al., 1997; Knoph Berg et al., 2011 Nunes et al., 2014; Tiller & Treasure, 1998). Results of the current study support this finding, with more than half of the women identified with possible disordered eating by the DEAPS reporting no history of diagnosed or undiagnosed disordered eating symptomatology. Among those who screened positive on the DEAPS, disordered eating cognitions and attitudes were high and prevalent, more so than disordered eating behaviours, as indicated by examination of individual item frequencies on the DEAPS. This finding is expected given most of the DEAPS items assessed cognitive and affective ED symptomatology, thus a positive screen is indicative on increased

levels of such symptoms. This observation is also consistent with the findings of Easter et al. (2013). Collectively, both findings underscore the potential for women experiencing disordered eating symptomatology to be unidentified in antenatal care if screening is conducted selectively based on previous history and/or presenting symptomatology, as such factors may not by disclosed by a woman or be overt to a clinician.

Although disordered eating cognitions and attitudes were more prevalent among those who obtained a positive screening result on the DEAPS, nearly half of those who screened positive were engaging in behaviours to prevent pregnancy-related bodily changes such as dietary restriction, meal skipping, purging, and excessive exercise. As such, although pregnancy may have had a protective benefit in preventing behavioural manifestations of disordered eating for some women, this benefit only extended to half of the women who screened positive. This is an important and concerning finding given the undesirable maternal and fetal impacts associated with such behaviours (Bulik et al., 1999; Bulik et al., 2009; Fornari et al., 2014; Linna et al., 2014; Micali et al., 2007; NICE, 2017a; Solmi et al., 2013). However, as noted in Chapter 3, distress caused from cognitive and attitudinal symptoms could have detrimental and lasting impacts on an unborn child, depending on the timing of cortisol exposure (see Davis & Sandman, 2010, for a review). With high levels of weight and shape concern during pregnancy, these women are at considerable risk for the onset of behavioural manifestations in the postpartum period when their child no longer serves to be a protective function (Astrachan-Fletcher et al., 2008; Easter et al., 2015; Lemberg & Phillips, 1989). Notably, nearly half the women in the positive screening group reported a significant investment of time researching methods to rapidly lose weight in the postpartum period. This not only highlights the importance of ongoing screening into the postpartum period with an appropriate psychometric instrument, but also the possibility that early intervention or management in pregnancy may prevent or mitigate symptom exacerbation or progression

postnatally (Astrachan-Fletcher et al., 2008; Edelstein & King, 1992), particularly as pregnancy is believed to be a unique window to make significant and sustainable behavioural changes (Wiles, 1994).

Limitations / Future Research

While the current study has several strengths to consider including the development of pregnancy-specific disordered eating attitudes screening instrument and the ability to recruit a sufficiently large sample of women at various stages of pregnancy to test the instrument in a development and validation sample, several limitations must be noted. First, although the intention of the current study was to develop an instrument that measured disordered eating symptoms in pregnancy (inclusive of cognitive, behavioural, and affective components), the final instrument was more reflective of cognitive and affective ED symptoms (i.e., attitudes). Despite this, the instrument is likely to be useful given previous research has indicated behavioural symptoms of disordered eating are often low or reduced during pregnancy, yet ED cognitions are often high (Blais et al., 2000; Crow et al., 2008; Easter et al., 2013; Micali et al., 2007). As the reduction in behavioural manifestations of disordered eating is often attributed to the protective function the unborn child temporarily provides during pregnancy, which is subsequently removed following birth, the postpartum period is believed to be especially vulnerable for the onset or resurgence of active ED behaviours in women with high levels of weight and shape concern. As such, the DEAPS may be an apt measure to identify women experiencing disordered eating attitudes for monitoring/intervention prior to the risk of the postpartum period.

Second, the development sample was homogeneous particularly in terms of ethnicity (i.e., most participants identified as Caucasian), therefore future studies should examine the psychometric performance of the DEAPS with ethnically diverse samples to investigate conceptual equivalence (Geisinger, 2003). A growing body of literature suggests a similar

prevalence of disordered eating symptomatology among non-pregnant ethnic minority samples (Cachelin, Rebeck, Veisel, & Striegel-Moore, 2001; Gordon, Perez, & Joiner, 2002; Marques et al., 2011; Smolak & Striegel-Moore, 2001; Solmi, Hatch, Hotopf, Treasure, & Micali, 2014), therefore appropriate validation with other ethnicities in pregnancy is required to determine whether the instrument may be ethnocentrically biased (Solmi et al., 2015). Third, due to the voluntary nature of the study, it is possible some participants may have been more compliant in completing the questionnaires because of a pre-existing interest and/or motivation. Similarly, as the current study was completed online in a self-report format with no feedback, validation in an antenatal care setting where all women are administered the DEAPS is required to establish validity generalisation. Not only would this minimise the impact of participant attribute variables (e.g., motivation and interest), it would also allow the feasibility and ease of administration to be examined more authentically and the impact of clinician feedback following scoring to be assessed.

The absence of a true gold standard measure in pregnancy is likely to have an ongoing impact on validation efforts; however, as noted in Chapter 4, the EDE-Q is the most appropriate reference standard at the current time, not the EDE interview. Further research in this area is required to determine an appropriate method to validate the EDE-Q. Recent pregnancy-modifications to the EDE interview (see Kolko et al., 2017) also require additional research to determine if the pregnancy-modified EDE interview could be used as a gold standard in future validation studies. Future research should consider alternate validation standards such as determining disordered eating caseness from clinical interviews conducted by clinicians with expertise in EDs and obstetrics. The predictive validity of the DEAPS must also be evaluated to determine how useful the DEAPS is in predicting current or future negative consequences that may be associated with ED symptomatology in pregnancy (e.g.,

low birth weight, prematurity, miscarriage, increased need for caesarean section, maternal and/or infant physical and mental health conditions, or medical utilisation).

Lastly, although counterbalancing the questionnaires likely minimised any significant order effects, completing all three scales may still have impacted the results through priming, thereby contributing to favourable sensitivity and specificity estimates across the scales. As such, it would be beneficial for future research to randomly assign women to either the DEAPS and EDE-Q or the SCOFF and EDE-Q to compare the sensitivity and specificity estimates of the DEAPS and SCOFF independently. It may also be beneficial to include a second administration of each instrument within a short period (i.e., 14 days) to ascertain temporal stability estimates, in addition to including a measure of general psychopathology to assess for comorbidities and determine whether discriminant validity can be established.

Conclusion

Findings of the current study provide preliminary evidence that the DEAPS constitutes a reliable, valid, and user-friendly instrument to assess and screen for disordered eating attitudes during pregnancy. Specifically, the DEAPS demonstrated a high level of internal consistency, appropriate content validity, good construct validity, and strong concurrent criterion-related validity. Given that almost a quarter of the sample had possible symptoms of disordered eating, routine antenatal screening to identify and support women who experience such symptoms is vital. Future research should further elucidate the validity of the DEAPS in different samples, study designs, settings, and administration methods. In particular, evaluation of the DEAPS in a clinical setting is crucial to establish external validity and exploration of predictive validity is vital to determine clinical utility.

CHAPTER SIX

General Discussion

Chapter Overview

This thesis presented four sequential studies that were carried out with the overarching aim of understanding and improving the identification of disordered eating in pregnancy. This final chapter begins by summarising the key findings from these studies, with reference to the objectives stated in Chapter 1. The clinical implications of this research are then discussed. Lastly, the main strengths and limitations of the research project are considered, with suggestions for future research provided.

Summary of Findings

As noted in Chapter 1, the current study had four main research questions, which were addressed sequentially in the current thesis. To contextualise this chapter, these research questions are re-stated below:

1. How does disordered eating manifest in pregnancy? (Chapter 3)
    1.1. Is this similar or distinct from disordered eating in a non-pregnant context?
    1.2. Does this perception differ between experienced health professionals and women with a lived experience?
2. How is disordered eating symptomatology distinguished from pregnancy-appropriate symptomatology? (Chapter 3)
    2.1. Where is the threshold between the two constructs and how do experienced health professionals determine this distinction?
    2.2. Does this perception differ between experienced health professionals and women with a lived experience?
3. Should screening for disordered eating occur in antenatal care and, if so, should this occur on a universal or selective basis? (Chapter 3)

4. What instruments are currently available to screen for disordered eating in pregnancy? (Chapter 4)
    4.1. If available, are these tools psychometrically sound and validated for use in pregnancy? (Chapter 4)
    4.2. If not, could a standardised and psychometrically sound pregnancy-specific screening instrument for detecting disordered eating be developed? (Chapters 4 and 5)

Consensus on the Features and Assessment of Disordered Eating in Pregnancy

Studies 1 and 2 used the Delphi consensus technique to garner the collective opinion of two expert panels in relation to disordered eating in pregnancy. Study 1 recruited a panel of international clinicians and researchers (N = 26) with expertise in the field of disordered eating, particularly in relation to pregnancy and/or women's health (professional expertise). Study 2 recruited women (N = 15) who identified with a lived experience of disordered eating in pregnancy (consumer expertise). Both panels sought to determine whether consensus could be reached on: 1) the features of disordered eating in pregnancy, 2) factors that may assist clinicians to distinguish disordered eating from pregnancy-appropriate symptomatology, and 3) whether assessment of disordered eating should occur in antenatal care, and if so, under what circumstances and using which methods. Previous research had not systematically explored these questions.

Overall, there was a high level of agreement within and across the two panels, despite the differences in their expertise (Jorm, 2015). Both panels perceived disordered eating in pregnancy to be somewhat similar, yet also distinct, to the experience of disordered eating in a non-pregnancy context, answering the first research question of thesis. The term "disordered eating" was conceptualised as a multidimensional experience, inclusive of behavioural, physical, cognitive, and affective symptomatology. Several unique pregnancy-

specific disordered eating symptoms were endorsed across both panels including overvaluation of the offspring's weight and shape (e.g., desire for the baby to be "small" or "petite"), rationalisation of self-induced vomiting as pregnancy-appropriate, and emotional detachment from the pregnancy. Behaviours often normalised outside of pregnancy, such as the use of natural supplements (e.g., tea detoxes) for weight loss, were also considered to be reflective of disordered eating in pregnancy and cause for concern if disclosed to clinicians practicing in this area. While the exact delineation between disordered eating and pregnancy-appropriate symptomatology remained difficult to explicitly state at end of Study 1 and Study 2, various quantitative and qualitative factors to assist practitioners evaluate concerning symptoms were established. As such, the most appropriate answer to the second research question is that delineating the clinical overlap between pregnancy-appropriate symptomatology and disordered eating is specific to an individual and/or the situation. To assist with the process of distinction, clinicians can use a range of common practice clinical skills including evaluating the results of appropriate psychometric instruments, undertaking functional analysis, considering the impact and impairment associated with individual symptom presentations, identifying informational discrepancies, exploring and considering the history of the woman and her family, and utilising clinical judgment.

Antenatal screening of disordered eating was also perceived to be crucial by both panels, answering the third research question of this thesis. There was almost unanimous consensus in both panels that such screening needs to occur in a routine manner. Several researchers over the past decade have advocated for routine screening in antenatal care (Abraham, 2001; Franko & Spurrell, 2000; Harris, 2010; Leddy et al., 2009; Squires et al., 2014). Recent literature has also suggested that most women perceive mental health screening during pregnancy to be highly beneficial and feel most comfortable when antenatal practitioners initiate the screening process in a routine manner (see Kingston et al., 2015).

Despite agreement in the professional panel that various assessment methods would be relevant in assessing disordered eating in pregnancy, only a small number of these methods were considered appropriate by the consumer panel. Psychometrically sound brief screening instruments were favoured by both panels, perceived to be most feasible for clinicians to administer and most comfortable for women accessing antenatal care. It was emphasised, however, that any screening instrument must be pregnancy-specific and delivered in an authentic and caring manner to be beneficial. Consistent with previous literature (Blais et al., 2000; Easter et al., 2013; Koubaa et al., 2005; Patel et al., 2002), concerns regarding the validity of existing screening instruments were expressed, particularly use of the SCOFF questionnaire. This highlighted the need for a systematic review to determine whether any pre-existing measures of disordered eating were suitable for use in pregnancy.

Systematic Review of Disordered Eating Measures Validated in Pregnancy Samples

Accordingly, Study 3 aimed to identify and evaluate the performance (reliability and validity) of general measures of disordered eating in pregnancy samples. Of the sixteen instruments identified during full text review, only three self-report inventories (EDE-Q, EDI-2, and DEBS) and one semi-structured clinical interview (EDE) had some form of psychometric information available. Most studies reported reliability. Only two reported validity. No studies assessed screening accuracy (i.e., sensitivity and specificity). The psychometric properties of the instruments reported in the eight publications were evaluated using a standardised performance appraisal tool developed by Terwee et al. (2007). Two instruments did not receive any positive ratings using the Terwee et al. (2007) criteria, while two received a positive rating in only one domain. Of the four instruments assessed, the EDE-Q had the most psychometric information available, with preliminary evidence to suggest possible utility in pregnancy. Although the EDE interview is considered the pre-eminent

instrument in the field of EDs and the standard by which all other EDs instruments are validated (Berg et al., 2012), there was insufficient evidence to confirm the EDE is currently suitable for use in pregnancy. As such, the EDE-Q was considered the closest gold standard proxy for identifying disordered eating in pregnancy. Furthermore, despite empirical literature and various antenatal guidelines encouraging clinicians to screen for disordered eating using the SCOFF (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; NEDC, 2015), no evidence was found to support this recommendation.

Overall, the review was limited by the small number of studies included (N = 8), highlighting the significant paucity of research investigating accurate and reliable screening/measurement of disordered eating symptomatology in pregnancy. To partially answer the fourth research question of this thesis, other than preliminary evidence for the EDE-Q, there was little to no evidence to recommend or support that general measures of disordered eating are appropriate or suitable for use in pregnancy. There was also a strong need for research exploring the validity of existing self-report inventories in pregnancy, including the EDE-Q. Without reliable and valid measures of disordered eating in pregnancy, researchers and clinicians will have difficulty identifying, measuring, and monitoring disordered eating symptoms in pregnancy. As such, development of a pregnancy-specific instrument was required.

Development and Evaluation of a Pregnancy-Specific Screening Instrument

To address the limitations of existing measures of disordered eating, Study 4 aimed to develop and evaluate an instrument that could identify symptoms of disordered eating in pregnancy. A brief, inexpensive, straightforward, and psychometrically sound instrument was desired. A six-step instrument development approach was utilised, with item content derived from the symptom attributes that reached consensus across both panels in Studies 1 and 2

(i.e., the Delphi studies). Following a two-stage instrument refinement process, the final instrument consisted of 14 self-report dichotomous items (i.e., yes/no). As the final instrument items were more reflective of cognitive and affective symptomatology of disordered eating, the instrument was labeled the Disordered Eating Attitudes in Pregnancy Scale (DEAPS). Overall, the DEAPS demonstrated a high level of internal consistency, appropriate content validity, good construct validity, and very strong concurrent criterion-related validity, thereby answering the fourth research question of this thesis. These psychometric properties were also consistent across the primary development sample (n = 222) and cross-validation sample (n = 222), suggesting the strong psychometric properties of the DEAPS were not distorted by chance effects (DeVellis, 2012). This provided preliminary evidence for the stability and robustness of the DEAPS.

At a cut-off score of eight, the DEAPS produced an overall sensitivity of 93.2 percent, specificity of 84.2 percent, and an AUC value of .95, indicating the DEAPS had excellent accuracy in correctly identifying cases and non-cases of disordered eating in the primary development and cross-validation samples. In comparison to the DEAPS, the SCOFF had an overall sensitivity of 77.2 percent and specificity of 83.8 percent. While these sensitivity and specificity estimates are consistent with the performance of the SCOFF in a non-pregnant context (see Botella et al., 2013, for a review), the DEAPS outperformed the SCOFF when identifying symptoms of disordered eating in the study. Notably, the SCOFF did not identify 22.8 percent of women with disordered eating symptoms, compared to 6.8 percent in the DEAPS. This difference in false negatives was considerable given the specificity of the DEAPS was also high. With only 14 items and taking less than two minutes to administer, the DEAPS has a similar level of brevity and ease of administration to the SCOFF (one of its main advantages), with a greater level of accuracy (see Chapter 5). Participants also reported

a similar level of acceptability and comfort when completing the SCOFF and the DEAPS, compared to the EDE-Q, which was perceived to be less comfortable to complete.

While estimating the prevalence of disordered eating in the sample was not a primary aim of Study 4, almost a quarter of the sample displayed disordered eating attitudes. This highlighted the importance and need for routine antenatal screening to identify and support women who experience such symptoms prior to the postpartum period. With further evaluation to determine robustness and predictive validity, particularly in a clinical setting, antenatal screening of disordered eating attitudes may be feasible via administration of the DEAPS.

Implications and Recommendations

The findings presented in Chapters 3, 4, and 5 have implications for clinicians, policy makers, and researchers. Specific implications are detailed in each relevant chapter. This section highlights and summarises some of the key implications.

Implications for Perinatal Mental Health Screening

As noted in Chapter 1, Australia's focus on perinatal mental health is strongly linked to enacting and achieving two of the millennium development goals outlined by the UN: improving and enhancing maternal health and reducing child mortality (UN, 2014; WHO, 2009). The poor rate at which mental health conditions are identified during the perinatal period is a key barrier to effectively supporting women during this life stage (Bowen et al., 2012; Coates et al., 2004; Spitzer et al., 2000). Over the past decade, perinatal mental health efforts have largely focused on implementing universal screening programs for depression, anxiety, and psychosocial factors that affect a woman's mental health (Austin et al., 2017). Australia has been a global leader in this field, with the federal government recently announcing that from November 2017 women will have access to Medicare-funded mental health screening during pregnancy and up to two months postpartum (Australian Government

Department of Health, 2017). While this is a positive and promising change in supporting women's mental health during the perinatal period, such screening focuses on depression, anxiety, domestic violence, and substance use. Screening for disordered eating continues to receive little attention, despite previous research suggesting up to 27.8 percent of women (1 in 4) experiencing such symptoms during pregnancy (Broussard, 2012). Results of Chapter 5 were consistent with this research, revealing almost a quarter of the sample (22%) screened positive for disordered eating symptomatology. This finding supports the suggestion that disordered eating is occurring at a similar rate to other common mental health conditions during pregnancy (e.g., depression and anxiety), and thus represents an important maternal health concern that warrants consideration by policy makers for inclusion in perinatal mental health clinical guidelines.

Identifying, monitoring, and supporting women with disordered eating symptomatology during pregnancy is crucial as such symptoms have been linked to several negative consequences such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and other obstetric and postpartum difficulties (Linna et al., 2014; Watson et al., 2014). Taking into consideration these potential consequences and the results of section three in the professional and consumer Delphi studies (see Chapter 3), the most significant clinical implication from this thesis is the need to integrate universal or routine screening of disordered eating symptomatology into antenatal and/or obstetric care to allow the provision of services and support in the most acceptable and accessible manner. This suggestion is also consistent with assertions made by previous researchers and advocacy organisations, globally (Abraham, 1998; Austin et al., 2017; Barnett et al., 2005; beyondblue, 2008; Buist et al., 2005; Franko & Spurrell, 2000; Harris, 2010; Kingston et al., 2015; Lowes et al., 2012; Mitchell & Bulik, 2006).

As noted by the National Eating Disorders Collaboration (2015), there are several screening opportunities for clinicians during maternity care including the initial pregnancy consultation (i.e., confirmation of pregnancy), various ultrasound appointments (particularly 12- and 20-weeks), the prenatal hospital admission interview, and third trimester check-ups. Each of these scenarios provides the opportunity for early detection, potentially increasing the likelihood of women receiving additional support during pregnancy, which may have protective effects for the mother and her offspring (Fornari et al., 2014). The consumers in Study 2 (Chapter 3) reported that routine screening would be best implemented using a brief screening instrument that is easy to administer, understand, score, and interpret. The excellent sensitivity and specificity of the DEAPS revealed in Study 4 (Chapter 5) suggests this new instrument may be ideal and more accurate than the SCOFF, which had a higher rate of false negatives, meaning that women with disordered eating symptomatology were not being accurately identified by the SCOFF. This is an important finding that should be considered by researchers, clinicians, and policy makers given the widespread endorsement of the SCOFF in research and clinical guidelines as an appropriate screening instrument in pregnancy (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; NEDC, 2015). Additional psychometric examination of both the DEAPS and SCOFF is, however, required prior to any explicit instrument recommendation (see Limitations section of this chapter for further details).

Some authors have argued that routine screening may cause undue psychological harm to unaffected women or deplete valuable resources and time, thereby favouring selective or indicated screening (Rollans et al., 2013; Shakespeare et al., 2003). This concern has mostly arisen from a small number of qualitative studies in which some women have reported negative experiences with perinatal mental health screening (Rollans et al., 2013; Shakespeare et al., 2003); however, most studies report general acceptability by women

(Buist et al., 2006; Gemmil, Leigh, Ericksen, & Milgrom, 2006; Leigh & Milgrom, 2007; Matthey et al., 2005). Results of the current thesis supported the general acceptability of perinatal mental health screening, with the women in Study 4 (Chapter 5) reporting a relatively high level of acceptability and comfort completing the disordered eating screening tools examined. Recent literature has also indicated most women perceive mental health screening during pregnancy to be highly beneficial and feel most comfortable when antenatal practitioners initiate the screening process in a routine manner (see Kingston et al., 2015). At a minimum, routine screening of a woman's eating-related behaviours, attitudes, and thoughts opens a dialogue between a woman and her antenatal practitioner about such concerns, which can often be difficult to approach and may facilitate further symptom disclosure, reduce stigma, and enhance the therapeutic relationship. It is, however, vital that clinicians approach screening in sensitive and non-judgmental manner.

Study 4 (Chapter 5) indicated that clinicians working in antenatal care must be vigilant of disordered eating symptomatology, regardless of sociodemographic characteristics and women's obstetric histories. Adding complexity to this, the professionals and consumers in the Delphi studies (Chapter 3) emphasised that disordered eating in pregnancy is multifaceted, encompassing behavioural, physical, cognitive, and affective manifestations. Accordingly, clinicians must be cognisant that an absence of physical or behavioural symptomatology alone does not necessarily imply a woman is unaffected by disordered eating attitudes during pregnancy. High levels of weight and shape concern (cognitions) often persist during pregnancy, despite a reduction in the behavioural components of disordered eating (Blais et al., 2000; Crow et al., 2008; Easter et al., 2013; Micali et al., 2007). A similar pattern was observed in Study 4 (Chapter 5), with cognitive and attitudinal based items most commonly endorsed by women who screened positive for possible disordered eating symptomatology. Behavioural symptomatology was, however, particularly prevalent for

binge eating behaviours (78%) among those who obtained a positive screen on the DEAPS. Behaviours to prevent pregnancy-related changes to body weight, shape, and size were also reported by almost half the women who screened positive. As such, it is important for clinicians to be aware that disordered eating in pregnancy reflects a spectrum of symptoms that do not necessarily result in physical weight or shape changes, and that particular exploration of binge eating behaviours and cognitions may be justified. These findings support previous work (Bulik et al., 2007; Knoph Berg et al., 2011; Soares et al., 2009; Watson et al., 2014).

The Delphi studies in Chapter 3 also revealed that disordered eating in pregnancy is likely to entail symptoms representative of the core characteristics of disordered eating (primary symptoms), in addition to secondary features, such as suicidal ideation, that are potentially reflective of various conditions in the broader mental health context in pregnancy. Practically, this may suggest that psychological screening instruments that assess symptomatology shared across a range of mental health conditions (i.e., a transdiagnostic framework) could act as platform or indicator to conduct further differential screening. Identifying or developing valid instruments that measure underlying cognitive behavioural processes in perinatal mental health may be a relevant topic for future research.

Similarly, if disordered eating is identified alongside symptomatology reflective of broader mental health concerns, additional screening should occur to rule out comorbid mental health issues. Future research may also wish to explore the comorbidity between disordered eating, other prevalent perinatal mental health conditions (e.g., depression and anxiety), and the impact this comorbidity has on the mother-child relationship. If a consistent pattern of comorbidity is revealed, it may be possible not only to assess women for specific cognitive or behavioural processes that increase vulnerability to a range of perinatal mental health conditions, but also to develop and deliver interventions that directly target these

underpinning vulnerabilities. This approach would be consistent with a transdiagnostic framework (Harvey et al., 1996; Harvey et al., 2004), which has been suggested for perinatal depression and anxiety (Coates, 2017; DeJong, Fox, & Stein, 2016; Poobalan et al., 2007). Notably, a transdiagnostic framework is used to understand EDs in a non-pregnant context and to treat the core mechanisms underpinning these conditions (see Fairburn et al., 2003). Research exploring whether these mechanisms also underpin the experience of disordered eating in pregnancy would be helpful, particularly if development of a pregnancy-specific ED intervention is required.

Implications for Management, Support, and Early Intervention

Results of the current thesis have highlighted the importance of routine screening for disordered eating in pregnancy for all women using a validated pregnancy-specific instrument. Routine psychosocial and mental health screening in the perinatal period is considered vital to increase the likelihood of women receiving early intervention and management, if needed (Austin et al., 2011; Austin et al., 2017; beyondblue, 2008). While the current thesis did not explore the process of managing and supporting women once a positive screening result was obtained, this is a vital area for future research. As noted in Chapter 1, the nature of early intervention following a positive screen would likely differ depending on the severity and frequency of symptoms and level of impairment. This may include regular monitoring and early education about sufficient and balanced eating to ensure a woman's caloric and nutrient intake is meeting the requirements of her own body and her child's (Chizawsky & Newton, 2006), preparing a woman for the range of physical changes that pregnancy entails (Andersen & Ryan, 2009; Czech-Szczapa et al., 2015), in addition to positively reinforcing maternal weight gain and shape changes by concurrently discussing foetal growth and development (Ward, 2008). To prevent the normalisation of disordered eating symptoms, it is vital that clinicians help women differentiate between symptoms of

disordered eating and changes in thoughts, feelings, and behaviours that occur because of pregnancy (Chizawsky & Newton, 2006). To provide and facilitate such support, more frequent and longer antenatal appointments may be required for women experiencing disordered eating in pregnancy (Harris, 2010; Lowes et al., 2012; NICE, 2004; Ward, 2008). This is potentially feasible under Australia's recent changes to the Medicare obstetric items or, alternatively, via the non-directive pregnancy support counselling items delivered by qualified mental health professionals (e.g., psychologists) under Medicare (Australian Government Department of Health, 2017).

In cases where there is risk of harm to the mother and/or unborn child, specialist multidisciplinary treatment incorporating medical monitoring, high-risk obstetric management, structured nutritional intervention, and psychotherapy may be necessary (Harris, 2010; Lowes et al., 2012; NEDC, 2015). As such, strong collaboration and communication between health care providers is essential and referral to a specialist ED service may be warranted to facilitate appropriate clinical intervention (Bulik et al., 2007; Lowes et al., 2012; NEDC, 2015). In such cases, care planning consistent with the NICE (2017a) guidelines on clinical management of antenatal mental health conditions is imperative to ensure a coordinated and transparent treatment approach.

Although there is clear guidance for monitoring and managing the physical risks associated with disordered eating in pregnancy, there is a paucity of literature regarding appropriate psychological intervention during this specific developmental stage (Crow et al., 2008; Soares et al., 2009; Tierney et al., 2013). For screening and early detection to be of greatest benefit, it is important that effective evidence-based interventions are identified (Public Health England, 2015; Wilson & Jungner, 1968). This is a key area for future research, including evaluation of existing interventions and/or the development of pregnancy-specific modifications, in addition to establishing clear referral pathways for clinicians.

Referral under the Better Access to Mental Health Care Initiative may be a viable option for women who require additional support from a psychologist. Early intervention during pregnancy could prevent progression into the postpartum period where symptoms are often exacerbated (Crow et al., 2008), in addition to mitigating or reducing undesirable foetal and maternal consequences that could have a lasting and detrimental impact (Fornari et al., 2014).

Implications for Professional Development and Education

Previous research has highlighted deficits in primary care physicians' and obstetric providers' knowledge of disordered eating symptomatology, which has been shown to affect both attitudes and clinical behaviour such as follow-up appointments and referral to specialist services for clinical assessment (Abraham, 2001; Currin, Waller, & Schmidt, 2009; Leddy et al., 2009; Morgan, 1999). Other research has revealed the low screening and assessment rates of disordered eating among antenatal practitioners (Abraham, 2001; Leddy et al., 2009; Morgan, 1999), mostly due to confidence concerns (Morgan, 1999). This highlights the prospective value of a brief screening instrument to initiate discussions about disordered eating symptomatology and potentially minimise the likelihood of subtle or concealed symptomatology going unnoticed. Moreover, it emphasises a need for greater education regarding disordered eating symptomatology and perinatal mental health in obstetric and midwifery training. The findings from Studies 1 and 2 (i.e., the Delphi studies) could encourage and assist the development of training resources to increase frontline antenatal health professionals' (e.g., obstetricians, GPs, midwives, and nurses) awareness, knowledge, and understanding of the expression and presentation of disordered eating in pregnancy.

Training to improve confidence and competence in performing sensitive perinatal mental health assessments is also required (Highet & Purtell, 2012; Reilly et al., 2014). It is likely that training in the administration, scoring, and interpretation of a pregnancy-specific screening instrument (such as the DEAPS developed in the current project) would be

relatively straightforward and could be achieved via self-directed (online) or guided (face-to-face) continuing professional development. Ultimately, professional development that incorporates training in both areas (i.e., symptom awareness/understanding and suitable instruments for symptom identification), in addition to increasing antenatal practitioners' skills in supporting women with disordered eating during pregnancy (practically or via referral pathways), will achieve the most integrated and beneficial outcome for women and the healthcare system.

Implications for Public Awareness

Popular media has traditionally labeled presentations of disordered eating in pregnancy as 'pregorexia'. This term is intended to describe women with an excessive fear of pregnancy-related weight gain who engage in various compensatory behaviours to avoid weight or shape changes that are characteristic of a healthy pregnancy (Hall-Flavin, 2015; Mathieu, 2009; Wallace, 2013). While this may be the experience of disordered eating for some women, it is not a universal experience. This thesis has highlighted that disordered eating in a spectrum of symptoms that can be expressed or experienced in numerous presentations. Cognitive or attitudinal based symptomatology appears to be the most common symptom expression. As the general population is increasingly reliant on popular media sources to obtain important information regarding their health and wellbeing (Fox & Duggan, 2013; Hogue et al., 2012), it is plausible that women experiencing symptoms inconsistent with the explanation of pregorexia may dismiss or downplay their symptoms. Clinicians (e.g., midwives, nurses, GPs, obstetricians, dietitians, and psychologists) interacting with pregnant women must be aware of the potential inaccuracies popular media presentations of disordered eating may result in and the need for appropriate psychoeducation to foster awareness and insight. It is also vital that popular media outlets disseminate accurate depictions of disordered eating in pregnancy to the general population to increase awareness and reduce

stigma around such symptoms, which may not be visible to a woman's social support network.

Implications for Researchers

In addition to the screening, intervention, and educational implications derived from the current thesis, there are several important research implications worth noting, mostly related to the measurement of disordered eating in research. To improve estimates of disordered eating in pregnancy and facilitate comparison across studies, achieving some form of consistency in measurement is crucial. This not only refers to the instrument utilised, but also the timing of administration and cut-off employed to determine clinical significance. As noted in Chapter 4, it is imperative that measurement of disordered eating during pregnancy is achieved using instruments or scales validated for this context. At the current time, there is an absence of research investigating accurate and reliable screening/measurement of disordered eating symptomatology in pregnancy. As such, there is a strong need for research exploring the validity of existing self-report inventories in pregnancy, including the EDE-Q and the SCOFF, and the validity of recent pregnancy-modifications to the EDE interview. This research is vital in identifying the most appropriate gold-standard instrument for comparison when new instruments such as the DEAPS are developed. The current thesis provides compelling preliminary evidence for the utility of the DEAPS in screening and detecting disordered eating symptomatology; however, further research is required to establish validity in different samples, settings, and geographic locations. Analysis of the DEAPS using Rasch modeling may also be beneficial if further item reduction is desired to create an even briefer screening instrument.

Limitations of the Research Program

Study-specific strengths and limitations have been highlighted and discussed throughout this thesis. As such, this section presents a general overview of main issues pertaining to the program of research and opportunities for future research.

First, while the two Delphi studies outlined in Chapter 3 provided a preliminary expert-derived template for understanding and distinguishing disordered eating from pregnancy-appropriate symptomatology, it is acknowledged that the list of symptom attributes and delineating foci generated may not have been exhaustive. In other words, additional or unique symptoms of disordered eating may also exist. As this expert-derived template was used to develop the Disordered Eating Attitudes in Pregnancy Scale (DEAPS), it is possible the instrument may not have complete construct coverage; however, given the instrument was developed from the results of two independent expert panels and was then subjected to a thorough piloting process by consumers and other health care professionals, the probability that important construct dimensions were missed is less likely. In a clinical sense, it is not unreasonable to suggest that clinical judgment should occur alongside administration of any psychometric instrument. Furthermore, it is acknowledged that the conceptualisation of disordered eating in current PhD, which focused on subclinical ED symptoms, is not the only conceptualisation of disordered eating that exists (e.g., external eating, disinhibited eating, emotional eating) and these forms may be more pertinent when identifying binge eating and obesity-related behaviours such as loss of control over eating (Hou et al., 2011). Such conceptualisations may also be more relevant when assessing disordered eating in multiethnic samples (Solmi et al., 2015).

Second, although the present research aimed to delineate disordered eating from pregnancy-appropriate abnormal eating, it became clear early in the Delphi studies that this aim was unattainable due to the heterogeneous presentation of disordered eating and

individual differences between women (current and historical factors). Furthermore, it was neither feasible nor helpful to identify thresholds and/or strict criteria for every possible symptom, as this would move toward a disorder approach and the focus of the current study was disordered eating symptomatology, not clinical eating disorders. While results of the current study did not entirely clarify the nuanced distinction between disordered eating and pregnancy-appropriate symptomatology, both panels endorsed various quantitative and qualitative factors to assist practitioners evaluate concerning symptoms.

Third, the absence of a true gold standard for the measurement/assessment of disordered eating symptomatology in pregnancy meant that validating the DEAPS was challenging. When a new instrument is developed, particularly a screening test, it must be benchmarked against an agreed upon 'gold standard' test (Greenhalgh, 1997). As noted in Chapter 5, a true gold standard is often unavailable, therefore, an 'alloy gold' standard or a proxy gold standard is recommended (Troy et al., 1996). The systematic review in Chapter 4 revealed the EDE-Q is currently the closest gold standard proxy, over and above the EDE interview, which has traditionally been used as the gold standard comparison instrument in ED research (Berg et al., 2012). Therefore, the DEAPS was validated against the EDE-Q, rather than the EDE; however, there is only preliminary evidence to support the use of the EDE-Q in pregnancy. This may have affected the evaluation of criterion-related validity in Chapter 5. It is also acknowledged that self-report measures generally result in higher prevalence rates than interviews (Paulson & Bazemore, 2010), which may affect the sensitivity and specificity estimates of the DEAPS. It could, however, be argued that the low-risk nature of self-report instruments means that individuals are more likely to disclose symptoms due to reduced fear of stigma, resulting in more accurate prevalence representations. As noted repeatedly throughout this thesis, it is well documented that fear of stigma and negative attitudes significantly hampers disclosure of mental health concerns in

perinatal care (Franko & Spurrell, 2000; Franko & Walton, 1993; Highet et al., 2014; Hollifield & Hobdy, 1990; Morgan, 1997; Newton & Chizawsky, 2006; Tierney et al., 2013).

Fourth, it is acknowledged that the psychometric properties of the DEAPS are preliminary and further validation in large samples is crucial prior to determining clinical utility. In particular, the predictive validity of the DEAPS must be evaluated to determine the usefulness of the DEAPS in predicting current or future negative consequences that may be associated with disordered eating symptomatology in pregnancy (e.g., miscarriage, prematurity, low birth weight, increased need for caesarean section, maternal and/or infant physical and mental health conditions, or medical utilisation). It is also noted that the validation sample in Chapter 5 was self-selected and relatively homogenous, particularly in terms of ethnicity (e.g., most of the women were Caucasian), which may have resulted in culturally biased instrument. To increase validity generalisation, the psychometric performance of the DEAPS must be assessed with ethnically diverse samples. Similarly, as the DEAPS was completed in an anonymous online self-report format, validation in an antenatal care setting where all women are administered the DEAPS is imperative to establish external validity. As noted previously, validation in a clinical context would allow the feasibility and ease of administration to be examined more authentically, in addition to minimising the impact of selection and sample bias.

Lastly, it is recognised this thesis may be perceived as contributing toward the over-medicalisation of maternity care, with the WHO (2016) guidelines for a positive pregnancy experience reporting that traditional antenatal care has focused too heavily on clinical assessment and intervention. It is important to note, however, that preventative screening in antenatal care has traditionally focused heavily on physical health conditions often entailing invasive, disruptive, and/or uncomfortable medical procedures (e.g., amniocentesis, glucose testing, etc) that screen for important, but often low prevalence conditions (see Miller et al.,

2016). Mental health screening is arguably less invasive, often completed via validated psychometric instruments in a self-report format. Furthermore, recent literature has suggested most women perceive routine mental health screening during pregnancy to have high benefit and low harm (see Kingston et al., 2015), with less than four percent refusing or reporting discomfort (Austin et al., 2010; Chew-Graham et al., 2009; Miller et al., 2009). The WHO (2016) guidelines for a positive pregnancy experience have also indicated that women desire additional support coping with psychosocial difficulties during pregnancy and the transition to motherhood. Finally, although routine screening may be perceived as time consuming and resource intensive, health professionals practicing in settings where routine psychosocial assessment and perinatal mental health screening has been implemented have found this approach to be both feasible and effective in identifying perinatal mental health conditions (Flynn et al., 2010; Mitchell & Coyne, 2009; Reay et al., 2011; Sword et al., 2008) and building rapport with women if delivered in a sensitive, woman-centered manner (Rao et al., 2007).

Despite these limitations, this program of research has highlighted some key issues in the identification of disordered eating in pregnancy and allowed a series of important implications to be derived. One of the main strengths of this thesis was the logical methodology that allowed the findings of the two Delphi studies to inform the instrument development process, contributing to the robust psychometric properties of the DEAPS. The large sample of pregnant women recruited in Chapter 5 also enabled the data set to split into primary development and cross-validation subsamples to assess validity shrinkage and reduce the impact of chance effects. Lastly, the systematic review in Chapter 4 may be useful for clinicians or researchers wanting to know whether general measures of disordered eating, conceptualised as subclinical ED symptoms, are appropriate for use in pregnancy.

Conclusion

Disordered eating presents a range of risks to the health and psychological wellbeing of women and their children from conception through to the postnatal period. Despite affecting women during their prime reproductive years, it has often been erroneously assumed that pregnancy functioned as a reprieve from eating and body image related disturbance. Research over the past decade has revealed that pregnancy alone does not assure remission or protection from disordered eating symptomatology and, in some cases, may trigger the onset of such symptoms. Disordered eating is, however, often undetected and undisclosed in pregnancy due to fear of stigma, poor knowledge and awareness of such symptoms during this period, difficulty distinguishing disordered eating from normative pregnancy symptoms, and potentially unsuitable assessment instruments. Importantly, antenatal providers are well positioned to screen for and identify disordered eating concerns as it is one of the rare occurrences in which women are heavily engaged in systematic and consistent healthcare, with various screening opportunities. The overarching aim of this thesis was to improve the identification of disordered eating in pregnancy.

Overall, this thesis revealed that disordered eating is a relatively common experience during pregnancy, affecting almost a quarter of women, and that routine/universal screening for such symptoms (particularly covert attitudes) is needed in antenatal care, similar to screening for antenatal depression and anxiety. This thesis has also provided preliminary evidence that implementation of universal screening may be feasible using the DEAPS, with strong psychometric properties and a high level of participant acceptability revealed. While further research is required to confirm the psychometric properties of the DEAPS in additional samples and different settings, this thesis has highlighted a significant need for policy makers to consider screening for disordered eating in perinatal mental health guidelines and the importance of clinician's being educated and aware of such symptoms.

Routine screening of disordered eating in pregnancy may facilitate early identification and management, contributing to a positive pregnancy experience and potentially mitigating associated long-term health consequences for women and children. Ongoing research in this area is vital, particularly the development and evaluation of evidence-based interventions to support women with disordered eating symptoms during pregnancy. It is hoped the current thesis provides the impetus for such work.

References

Aardoom, J. J., Dingemans, A. E., Slof Op't Landt, M. C. T., & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviours, 13,* 305-309. doi:10.1016/j.eatbeh.2012.09.002

Abe, C., Minami, J., Ohrui, M., Ishimitsu, T., & Matsuoka, H. (2007). Lower birth weight is associated with higher resting heart rate during boyhood. *Hypertension Research, 30*, 945–950. doi:10.1291/hypres.30.945

Abebe, D.S., Lien, L., & von Soest, T. (2012). The development of bulimic symptoms from adolescence to young adulthood in females and males: A population-based longitudinal cohort study. *International Journal of Eating Disorders, 45*, 737–745. doi:10.1002/eat.20950

Abraham, S. (2001). Obstetricians and maternal body weight and eating disorders during pregnancy. *Journal of Psychosomatic Obstetrics and Gynaecology, 22*, 159-163. doi:10.3109/01674820109049968

Abraham, S., Brown, T., Boyd, C., Luscombe, G., & Russell, J. (2006). Quality of life: Eating disorders. *Australian New Zealand Journal of Psychiatry, 40*, 150. doi:10.1111/j.1440-1614.2006.01762.x

Abraham, S., King, W., & Llewellyn-Jones, D. (1994). Attitudes to body weight, weight gain and eating behavior in pregnancy. *Journal of Psychosomatic Obstetrics and Gynaecology, 15*, 189–195. doi:10.3109/01674829409025645

Addington, D. E., McKenzie, E., Norman, R., Wang, J., & Bond, G. R. (2013). Essential evidence based components of first-episode psychosis services. *Psychiatric Services, 64*, 452-457. doi:10.1176/appi.ps.201200156.

Agras, W. S. (2001). The consequences and costs of the eating disorders. *Psychiatric Clinics of North America, 24,* 371-379. doi:10.1016/S0193-953X(05)70232-X

Akins R. B., Tolson, H. & Cole, B. R. (2005). Stability of response characteristics of a Delphi panel: Application of bootstrap data expansion. *BMC Medical Research Methodology, 5*, 37. doi:10.1186/1471-2288-5-37

Alderdice, F., Ayers, S., Darwin, Z., Greenwood, J., Joeemn, J., Kenyon, S., . . . & Walsh, J. (2013). Measuring psychological health in the perinatal period: Workshop consensus statement. *Journal of Reproductive and Infant Psychology, 31,* 431-438. doi:10.1080/02646838.2013.835039

Alexander, J., & Treasure, J. (2012). *A collaborative approach to eating disorders*. New York, NY: Routledge.

Allison, K. C., Wrotniak, B. H., Paré, E., & Sarwer, D. B. (2012). Psychosocial characteristics and gestational weight change among overweight, African American pregnant women. *Obstetrics and Gynecology International, 2012,* 1-9. doi: 10.1155/2012/878607

Allen, K. L., Byrne, S. M., Lampard, A., Watson, H., & Fursland, A. (2011). Confirmatory factor analysis of the Eating Disorder Examination-Questionnaire (EDE-Q). *Eating Behaviors, 12,* 143–151. doi:10.1016/j.eatbeh.2011.01.005

Allen, K. L., Byrne, D. M., Oddy, W. H., & Crosby, R. D. (2013). DSM-IV-TR and DSM-5 eating disorders in adolescents: prevalence, stability, and psychosocial correlates in a population-based sample of male and female adolescents. *Journal of Abnormal Psychology, 122,* 720. 732. doi:10.1037/a0034004

American Educational Research Association., American Psychological Association., & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.

Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *Bulletin of the World Health Organisation.* Retrieved from: http://www.who.int/bulletin/volumes/86/4/07-050112/en/

Andersen, A. E., & Ryan, G. L. (2009). Eating disorders in the obstetric and gynecologic patient population. *Obstetrics and Gynecology, 114,* 1353-1367. doi:10.1097/AOG.0b013e3181c070f9

Annagür, B. B., Kerimoglu, Ö. S., Gündüz, S., & Tazegül, A. (2014). Are there any differences in psychiatric symptoms and eating attitudes between pregnant women with hyperemesis gravidarum and healthy pregnant women? *Journal of Obstetrics and Gynecology Research, 40,* 1009-1014. doi:10.1111/jog.12274

Antonow-Schlorke, I., Schwab, M., Cox, L. A., Li, C., Stuchlik, K., Witte, O. W., Nathanielsz, P. W., McDonald, T. J. (2011). Vulnerability for the fetal primate brain to moderate reduction in global nutrient availability. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 3011-3016. doi:10.1073.pnas.1009838108

Astrachan-Fletcher, E., Veldhuis, C., Lively, N., Fowler, C., & Marcks, B. (2008). The reciprocal effects of eating disorders and the postpartum period: A review of the literature and recommendations for clinical care. *Journal of Women's Health, 17,* 227-239. doi:10.1089/jwh.2007.0550

Atkins, R. B., Tolson, H., & Cole, B. R. (2005). Stability of response characteristics of a Delphi bootstrap data expansion. *BMC Medical Research Methodology, 5,* 37-47. doi:10.1186/1471-2288-5-37

Austin, S.B. (2000). Prevention research in eating disorders: Theory and new directions. *Psychological Medicine, 30,* 1249-1262. doi:10.1017/s0033291799002573

Austin M-P., Highet, N., and the Guidelines Expert Working Group (2011). *Clinical practice guidelines for depression and related disorders – anxiety, bipolar disorder and puerperal psychosis – in the perinatal period. A guideline for primary care health professionals*. Melbourne: beyondblue.

Austin M-P., Highet, N., and the Expert Working Group. (2017). *Mental health care in the perinatal period: Australian clinical practice guideline.* Melbourne: Centre of Perinatal Excellence.

Austin, M-P., Kildea, S., & Sullivan, E. (2007). Maternal mortality and psychiatric morbidity in the perinatal period: Challenges and opportunities for prevention in the Australian setting. *Medical Journal of Australia, 186*, 364–367.

Austin, M-P., & Priest, S. R. (2004). New developments in perinatal mental health. *Acta Psychiatrica Scandinavica, 110*, 321-322. doi:10.1111/j.1600-0447.2004.00414.x

Australian Bureau of Statistics. (2008). *The 2007 National Survey of Mental Health and Wellbeing: Summary of results.* Canberra: ABS.

Australian Government Department of Health and Ageing. (2012). *Clinical practice guidelines: Module I.* Canberra: Australian Department of Health and Ageing.

Australian Government Department of Health. (2017). *Changes to MBS items for obstetrics services.* Retrieved from: http://www.health.gov.au/internet/mbsonline/publishing.nsf/Content/298E6E8507E0843ACA258146002613CB/$File/MBS%20Items%20Obstetric%20Services_23Oct2017.pdf

Australian Institute of Health and Welfare. (2012). *Experience of perinatal depression: Data from the 2010 Australian National Infant Feeding Survey*. Canberra: AIHW.

Bahk, J., Yun, S-C., Kim, Y-M., & Khang, Y-H. (2015). Impact of unintended pregnancy on maternal mental health: A causal analysis using follow up data of the Panel Study on Korean Children (PSKC). *BMC Pregnancy & Childbirth, 15,* 85. doi:10.1186/s12884-015-0505-4

Baker, C. W., Carter, A. S., Cohen, L. R., & Brownell, K. D. (1999). Eating attitudes and behaviours in pregnancy and postpartum: Global stability versus specific transitions. *Annals of Behavioural Medicine, 21*, 143–148. doi:10.1007/BF02908295

Baker, J., Lovell, K., & Harris, N. (2006). How expert are the experts? An exploration of the concept of ‘expert' within Delphi panel techniques. *Nurse Researcher, 14*, 59–70. doi:10.7748/nr2006.10.14.1.59.c6010

Bannatyne, A., & Stapleton, P. (2015). Educating medical students about anorexia nervosa: A potential method for reducing the volitional stigma associated with the disorder. *Eating Disorders: The Journal of Treatment and Prevention, 2*, 115-133. doi:10.1080/10640266.2014.976102

Bansil, P., Kuklina, E. V., Whiteman, M. K., Kourtis, A. P., Posner, S. F., Johnson, C. H., & Jamieson, D. J. (2008). Eating disorders among delivery hospitalisations: Prevalence and outcomes. *Journal of Women’s Health, 17,* 1523-1528. doi:10.1089/jwh.2007.0779

Bantel, K. A. (1993). Comprehensiveness of strategic planning: The importance of heterogeneity of a top team. *Psychological Reports, 73*, 35–49. doi:10.2466/pr0.1993.73.1.35

Bardone-Cone, A. M., & Agras, W. S. (2007). Psychometric properties of eating disorder instruments in black and white young women: Internal consistency, temporal stability, and validity. *Psychological Assessment, 19*, 356-362. doi:10.1037/10403590.19.3.356

Barker, E. D., Jaffee, S. R., Uher, R., & Maughan, B. (2011). The contribution of prenatal and postnatal maternal anxiety and depression to child maladjustment. *Depression and Anxiety, 28*, 696-702. doi:10.1002/da.20856

Barnes, J., Prescott, T., & Muncer, S. (2012). Confirmatory factor analysis for the Eating Disorder Examination Questionnaire: Evidence supporting a three-factor model. *Eating Behaviours, 13,* 379-381. doi:10.1016/j.eatbeh.2012.05.001

Baudet, M. L., Montastier, E., Mesthe, P., Oustric, S., Lepage, B., & Ritz, P. (2013). The SCOFF score: A screening tool for eating disorders in family practice. *e-SPEN Journal, 8*, 86-89. doi:10.1016/j.clnme.2013.03.001

Bauer, A., Parsonage, M., Knapp, M., Iemmi,V., & Adelaja, B. (2014). *The costs of perinatal mental health problems.* Centre for Mental Health and London School of Economics.

Bauer, A., Pawlby, S., Plant, D. T., King, D., Pariante, C. M., & Knapp, M. (2015). Perinatal depression and child development: Exploring the economic consequences from a South London cohort. *Psychological Medicine, 45,* 51-61 doi:10.1017/S003329 1714001044.

Bayley, T. M., Dye, L., Jones, S., DeBono, M., & Hill, A. J. (2002). Food cravings and aversions during pregnancy: Relationships with nausea and vomiting. *Appetite, 38*, 45-51. doi:10.1006/appe.2002.0470

Becker, G. E., & Roberts, T. (2009). Do we agree? Using a Delphi technique to develop consensus on skills of hand expression. *Journal of Human Lactation, 25,* 220–225. doi:10.1177/0890334409333679

Becker, A. E., Thomas, J. J., Bainivualiku, A., Richards, L., Navara, K., Roberts, A. L., Gilman, S. E., Striegel-Moore, R. H., & Healthy Fiji Study Group. (2010). Validity and reliability of a Fijian translation and adaptation of the Eating Disorder Examination Questionnaire. *International Journal of Eating Disorders, 43*, 171–178. doi:10.1002/eat.20675

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., . . . & Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Journal of the American Medical Association, 276*, 637-639. doi:10.1001/jama.276.8.637

Behar, R. A., González, J. A., Ariza, M. P., & Aguirre, A. S. (2008). Trastornos de la conducta alimentaria en mujeres embarazadas controladas en atención primar. *Revista Chilena de Obstetricia y Ginecología, 73*, 155-162.

Benton-Hardy, L. R., & Lock, J. (1998). Pregnancy and early parenthood: Factors in the development of anorexia nervosa? *International Journal of Eating Disorders, 24,* 223-226. doi:10.1002/(SICI)1098-108X(199809)24:2<223::AID-EAT12>3.0.CO;2-#

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2011). Psychometric evaluation of the eating disorder examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45*, 428–438. doi:10.1002/eat.20931

Berger U., Wick K., Hölling H., Schlack R., Bormann B., Brix C., …, & Strau B. (2011). Screening of disordered eating in 12-year-old girls and boys: Psychometric analysis of the German versions of SCOFF and EAT-26. *Psychotherapie, Psychosomatik & Medizinische Psychologie, 61*, 311–318. doi:10.1055/s-0031-1271786

Berk, L., Jorm, A. F., Kelly, C. M., Dodd, S., & Berk, M. (2011). Development of guidelines for caregivers of people with bipolar disorder: A Delphi expert consensus study. *Bipolar Disorders, 13*, 556–570. doi:10.1111/j.1399-5618.2011.00942.x

beyondblue. (2008). *Perinatal mental health national action plan 2008-2010: Full report*. Melbourne, VIC: beyondblue.

Biaggi, A., Conroy, S., Pawlby, S., & Pariante, C. M. (2016). Identifying the women at risk of antenatal anxiety and depression: A systematic review. *Journal of Affective Disorders, 191*, 62-77. doi:10.1016/j.jad.2015.11.014

Binford, R. B., Le Grange, D., & Jellar, C. C. (2005). Eating Disorders Examination versus Eating Disorders Examination-Questionnaire in adolescents with full and partial-syndrome bulimia nervosa and anorexia nervosa. *International Journal of Eating Disorders, 37,* 44-49. doi:10.1002/eat.20062

Bird, V. J., Le Boutillier, C., Leamy, M., Larden, J., Oades, G., Williams, L. J., & Slade, M. (2012). Assessing the strengths of mental health consumers: A systematic review. *Psychological Assessment, 24,* 1024-1033. doi:10.1037/a0028983

Birgegård, A., Norring, C., & Clinton, D. (2012). DSM-IV versus DSM-5: Implementation of proposed DSM-5 criteria in a large naturalistic database. *International Journal of Eating Disorders, 45,* 353-363. doi:10.1002/eat.20968

Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J., & Marteau, T. (1999). Consensus development methods: A review of best practice in creating clinical guidelines. *Journal of Health Services Research Policy, 4,* 236-248. doi:10.1177/135581969900400410

Blais, M. A., Becker, A. E., Burwell, R. A., Flores, A. T., Nussbaum, K. M., Greenwood, D. N., Ekeblad, E. R., & Herzog, D. B. (2000). Pregnancy: Outcome and impact on

symptomatology in a cohort of eating-disordered women. *International Journal of Eating Disorders, 27*, 140–149. doi:10.1002/(SICI)1098-108X(200003)

Blinder, B. J., Cumella, E. J., & Sanathara, V. A. (2006). Psychiatric comorbidities of female inpatients with eating disorders. *Psychosomatic Medicine, 68,* 454-462. doi:10.1097/01.psy.0000221254.77675.15

Boccardi, M., Bocchetta, M., Apostolova, L. G., Barnes, J., Bartzokis, G., Corbettea, G., . . . & EADC-ADNI Working Group (2015). Delphi definition of the EADC-ADNI harmonized protocol for hippocampal segmentation on magnetic resonance. *Alzheimer's & Dementia, 11*, 126–138. doi:10.1016/j.jalz.2014.02.009

Bond, K. S., Jorm, A. F., Kelly, C. M., Kitchener, B. A., Morris, S. L., & Mason, R. J. (2017). Considerations when providing mental health first aid to an LGBTIQ person: A Delphi study. *Advances in Mental Health: Promotion, Prevention, and Early Intervention, 15,* 183-197. doi:10.1080/18387357.2017.1279017

Bondas, T., & Eriksson, K. (2001). Women's lived experiences of pregnancy: A tapestry of joy and suffering. *Qualitative Health Research, 11*, 824-840. doi:10.1177/10497 3201129119415

Boscaglia, N., Skouteris, H., & Wertheim, E. H. (2003). Changes in body image satisfaction during pregnancy: A comparison of high exercising and low exercising women. *Australian and New Zealand Journal of Obstetrics and Gynaecology, 43,* 42-45. doi:10.1046/j.0004-8666.2003.00016.x

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS One, 6,* e20476. doi:10.1371/journal.pone.0020476

Bovopoulous, N., Jorm, A. F., Bond, K. S., LaMontagne, A. D., Reavley, N., Kelly, C. M., Kitchener, B. A., & Martin, A. (2016). Providing mental health first aid in the workplace: A Delphi consensus study. *BMC Psychology, 4,* 41-52. doi:10.1186/S40359-016-0148-x

Bowen, A., Bowen, R., Butt, P., Rahman, K., & Muhajarine, N. (2012). Patterns of depression and treatment in pregnant and postpartum women. *Canadian Journal of Psychiatry, 57,* 161–167. doi:10.1177/070674371205700305

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54,* 106-148. doi:10.1111/j.1467-6494.1986.tb00391.x

Broussard, B. (2012). Psychological and behavioural traits associated with eating disorders and pregnancy: A pilot study. *Journal of Midwifery & Women's Health, 57,* 61-66. doi:10.1111/j.1542-2011.2011.00089

Brown, A., & Avery, A. (2012). Healthy weight management during pregnancy: What advice and information is being provided. *Journal of Human Nutrition and Dietetics, 25,* 378-387. doi:10.1111/j.1365-277X.2012.01231.x

Buist, A. (2014). Perinatal mental health: Identifying problems and managing medications. *Australian Family Physician, 43,* 182-185.

Buist, A., & Bilsztra, J. (2006). *The beyondblue National Postnatal Screening Program, Prevention and Early Intervention 2001–2005, Final Report. Vol 1: National Screening Program*. Melbourne, VIC: Beyondblue.

Buist, A., Condon, J., Brooks, J., Speelman, C., Milgrom, J., Hayes, B., . . . & Bilsztra, J. (2006). Acceptability of routine screening for perinatal depression. *Journal of Affective Disorders, 93,* 233–237. doi:10.1016/j.jad.2006.02.019

Bulik, C. M., Hoffman, E. R., Von Holle, A., Torgersen, L., Stoltenberg, C., & Reichborn-Kjennerud, T. (2010). Unplanned pregnancy in women with anorexia nervosa. *Obstetrics and Gynecology, 116*, 1136-1140. doi:10.1097/AOG.0b013e3181f7efdc

Bulik, C. M., Sullivan, P. F., Fear, J. L., Pickering, A., Dawn, A., & McCullen, M. (1999). Fertility and reproduction in women with anorexia nervosa: A controlled study. *Journal of Clinical Psychiatry, 60*, 130-135. doi:10.1097/AOG.0b013e3181f7efdc

Bulik, C. M., Von Holle, A., Hamer, R., Knoph Berg, C., Torgersen, L., Magnus, P., Stoltenberg, C., . . . & Reichborn-Kjennerud, T. (2007). Patterns of remission, continuation and incidence of broadly defined eating disorders during early pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Psychological Medicine, 37*, 1109–1118. doi:10.1017/S0033291707000724

Bulik, C. M., Von Holle, A., Siega-Riz, A. M., Torgersen, L., Lie, K. K., Hamer, R. M., . . . & Reichborn-Kjennerud, T. (2009). Birth outcomes in women with eating disorders in the Norwegian Mother and Child cohort study (MoBa). *International Journal of Eating Disorders, 42*, 9-18. doi:10.1002/eat.20578

Bulliard, J-L., & Chiolero, A. (2015). Screening and overdiagnosis: Public health implications. *Public Health Reviews, 36,* 8. doi:10.1186/s40985-015-0012-1

Bunevicius, A., Kusminskas, L., Pop, V. J., Pedersen, C. A., & Bunevicius, R. (2008). Screening for antenatal depression with the Edinburgh Depression Scale. *Journal of Psychosomatic Obstetrics and Gynecology, 30,* 238-243. doi:10.3109/01674820903230708

Burton, A. L., Abbott, M. J., Modini, M., & Touyz, S. (2016). Psychometric evaluation of self-report measures of binge eating symptoms and related psychopathology: A systematic review of the literature. *International Journal of Eating Disorders, 49,* 123-140. doi:10.1002/eat.22453

Butterfly Foundation. (2012). *Paying the price: The economic and social impact of eating disorders in Australia.* Crows Nest, NSW: Butterfly Foundation.

Butterfly Foundation. (2017). *Disordered eating*. Retrieved from: https://thebutterflyfoundation.org.au/understand-eating-disorders/disordered-eating/

Caamaño F., Aguiar M., López-Otero M. J., & Takkouche B. (2002). Validation of SCOFF questionnaire among preteenagers. *British Medical Journal, 325,* 755-756. doi:10.1136/bmj.325.7367.755

Cachelin, F. M., Rebeck, R., Veisel, C., & Striegel-Moore, R. H. (2001). Barriers to treat for eating disorders among ethnically diverse women. *International Journal of Eating Disorders, 30,* 269-278. doi:10.1002/eat.1084

Cairns, K. E., Yap, M. B. H., Reavley, N. J., & Jorm, A. F. (2015). Identifying prevention strategies for adolescents to reduce their risk of depression: A Delphi consensus study. *Journal of Affective Disorders, 183,* 229–238. doi:10.1016/j.jad.2015.05.019

Carin, M., Lundgren, I., & Bergbom, I. (2011). First time pregnant women's experiences in early pregnancy. *International Journal of Qualitative Studies on Health and Wellbeing, 6,* 5600-5611. doi:10.3402/qhw.v6i2.5600

Carter, F. A., McIntosh, V. V. W., Joyce, P. R., Frampton, C. M., & Bulik, C. M. (2003). Bulimia nervosa, childbirth, and psychopathology. *Journal of Psychosomatic Research, 55,* 357-361. doi:10.1016/S0022-3999(02)00641-4

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research, 1,* 245-276. doi:10.1207/s15327906mbr0102_10

Cavelti, M., Kvrgic, S., Beck., E-M., Kossowsky, J., & Vauth, R. (2012). Assessing recovery from schizophrenia as an individual process: A review of self-report instruments. *European Psychiatry, 27,* 19-32. doi:10.1016/j.eurpsy.2011.01.007

Celio, A. A., Wilfley, D. E., Crow, S. J., Mitchell, J., & Walsh, B. T. (2004). A comparison of the Binge Eating Scale, Questionnaire for Eating and Weight Patterns-Revised, and Eating Disorder Examination-Questionnaire with instructions from the Eating Disorder Examination in the assessment of binge eating disorder and its symptoms. *International Journal of Eating Disorders, 36,* 434–444. doi:10.1002/eat.20057

Chamay-Weber, C., Narring, F., & Michaud, P.A. (2005). Partial eating disorders among adolescents: A review. *Journal of Adolescent Health, 37,* 416-426. doi:10.1016/j.jadohealth.2004.09.014

Chew-Graham, C. A., Sharp, D., Chamberlain, E., Folkes, L., & Turner, K. M. (2009). Disclosure of symptoms of postnatal depression, the perspectives of health professionals and women: a qualitative study. *BMC Family Practice, 10,* 7. doi:10.1186/1471-2296-10-7

Chizawsky, L. L., & Newton, M. S. (2006). Eating disorders: Identification and treatment in obstetrical patients. *Nursing for Women's Health, 10,* 482-488. doi:10.1111/j.1552-6356.2006.0097.x

Choi, S. K., Kim, J. J., Park, Y. G., Ko, H. S., Park, I. Y., & Shin, J. C. (2012). The simplified Edinburgh Postnatal Depression Scale (EPDS) for antenatal depression: Is it a valid measure for pre-screening? *International Journal of Medical Sciences, 9,* 40-46. doi:10.7150/ijms.9.40

Chorwe-Sungani, G., & Chipps, J. (2017). A systematic review of screening instruments for depression for use in antenatal services in low resource settings. *BMC Psychiatry, 17,* 112. doi:10.1186/s12888-017-1273-7

Clark, M. & Ogden, J. (1999). The impact of pregnancy on eating behaviour and aspects of weight concern. *International Journal of Obesity, 23*, 18–24. doi:10.1038/sj.ijo.0800747

Clark, A., Skouteris, H., Wertheim, E. H., Paxton, S. J., & Milgrom, J. (2009). My baby body: A qualitative insight into women's body-related experiences and mood during pregnancy and the postpartum. *Journal of Reproductive and Infant Psychology, 27,* 330-345. doi:10.1080/02646830903190904

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in scale development. *Psychological Assessment, 7,* 309-319. doi:10.1037/1040-3590.7.3.309

Clausen, L., Rokkedal, K., & Rosenvinge, J. H. (2009). Validating the Eating Disorder Inventory (EDI-2) in two Danish samples: A comparison between female eating disorder patients and females from the general population. *European Eating Disorders Review, 17,* 462-467. doi:10.1002/erv.945

Coaley, K. (2014). *An introduction to psychological assessment and psychometrics* (2nd ed.) Thousand Oaks, CA: Sage Publications.

Coates, R. (2017). *Assessment of perinatal mental health problems* (doctoral dissertation). Retrieved from: https://pdfs.semanticscholar.org/5fd7/b4b180df620bb161163be70d819947339c4b.pdf

Coates, A. O., Schaefer, C. A., & Alexander, J. L. (2004). Detection of postpartum depression and anxiety in a large health plan. *Journal of Behavioural Health Services and Research, 31,* 117–133. doi:10.1007/BF02287376

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th ed.). New York: McGraw-Hill.

Coker, E. L., & Abraham, S. F. (2015). Body weight dissatisfaction before, during, and after pregnancy: A comparison of women with and without eating disorders. *Eating and Weight Disorders, 20,* 71-79. doi:10.1007/s40519-014-0133-4

Coker, E. L., Mitchell-Wong, L. A., & Abraham, S. F. (2013). Is pregnancy a trigger for recovery from an eating disorder? *Acta Obstetricia et Gynecologica Scandinavica, 92,* 1407–1413. doi:10.1111/aogs.12256

Cole, Z. D., Donohoe, H. M., & Stellefson, M. L. (2013). Internet-based Delphi research: Case based discussion. *Environmental Management, 51,* 511-523. doi: 10.1007/s00267-012-0005-5

Conrad, R., Schablewski, J., Schilling, G., & Liedtke, R. (2003). Worsening of symptoms of bulimia nervosa during pregnancy. *Psychosomatics, 44,* 76–78. doi:10.1176/appi.psy.44.1.76

Conti, J., Abraham, S., & Taylor, A. (1998). Eating behaviour and pregnancy outcome. *Journal of Psychosomatic Research, 44,* 465-477. doi:10.1016/S00223999(97)00271-7

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104. doi:10.1037/0021-9010.78.1.98

Cotton, M. A., Ball, C., & Robinson, P. (2003). Four simple questions can help screen for eating disorders. *Journal of General and Internal Medicine, 18,* 53–56. doi:10.1046/j.1525-1497.2003.20374.x

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. *British Journal of Psychiatry, 150*, 782-786. doi:10.1192/bjp.150.6.782

Crisp, J., Pelletier, D., Duffield, C., Adams, A., & Nagy, S. (1997). The Delphi method. *Nursing Research, 46,* 116–118. doi:10.1097/00006199-199703000-00010

Crow, S. J., Agras, W. S., Crosby, R., Halmi, K., & Mitchell, J. E. (2008). Eating disorder symptoms in pregnancy: A prospective study. *International Journal of Eating Disorders, 41*, 277–279. doi:10.1002/eat.20496

Crow, S. J., Keel, P. K., Thuras, P., & Mitchell, J. E. (2004). Bulimia symptoms and other risk behaviors during pregnancy in women with bulimia nervosa. *International Journal of Eating Disorders, 36*, 220–223. doi:10.1002/eat.20031

Crowe, M., Inder, M., & Porter, R. (2015). Conducting qualitative research in mental health: Thematic and content analyses. *Australian and New Zealand Journal of Psychiatry, 49*, 616-623. doi:10.1177/0004867415582053

Crystal, S. R., Bowen, D. J., & Bernstein, I. L. (1999). Morning sickness and salt intake, food cravings, and food aversions. *Physiological Behaviour, 67*, 181-187. doi:10.1016/S0031-9384(99)00055-4

Culbert, K. M., Racine, S. E., & Klump, K. L. (2015). Research review: What we have learned about the causes of eating disorders: A synthesis of sociocultural, psychological, and biological research. *Journal of Child Psychology and Psychiatry, 56*, 1141-1164. doi:10.1111/jcpp.12441

Cureton, E. E. (1957). The upper and lower twenty-seven percent rule. *Psychometrika, 22*, 293-296.

Currin, L., Waller, G., & Schmidt, U. (2009). Primary care physicians' knowledge of and attitudes toward the eating disorders: Do they affect clinical actions? *International Journal of Eating Disorders, 42*, 453–458. doi:10.1002/eat.20636

Czech-Szczapa, B., Szczapa, T., Merritt, T. A., Wysocki, J., Gadzinowski, T., & Drews, K. (2015). Disordered eating attitudes during pregnancy in mothers of newborns requiring Neonatal Intensive Care Unit admission: A case control study. *Journal of Maternal-Fetal & Neonatal Medicine, 28*, 1711-1715. doi:10.3109/14767058.2014.966675

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the user of experts. *Management Science, 9*, 3, 458-467.

Dalkey, N. C., & Rourke, D. L. (1972). Experimental assessment of Delphi procedures with group value judgments. In N. C. Dalkey, D. L. Rourke, R. Lewis, & D. Snyder (Eds.). *Studies in the quality of life: Delphi and decision-making* (pp. 55-83). Lexington, MA: Lexington Books.

Darcy, A. M., Hardy, K. K., Crosby, R. D., Lock, J., & Peebles, R. (2013). Factor structure of the Eating Disorder Examination Questionnaire (EDE-Q) in male and female college athletes. *Body Image, 10,* 399-405. doi:10.1016/j.bodyim.2013.01.008

Darvill, R., Skirton, H., & Farrand, P. (2010). Psychological factors that impact on women's experiences of first-time motherhood: A qualitative study of the transition. *Midwifery, 26,* 357-366. doi:10.1016/j.midw.2008.07.006

Davies, S. (2015). *Annual report of the Chief Medical Officer 2014: The health of the 51% women*. London: Department of Health. Retrieved from: www.gov.uk/government/publications/chief-medical- o cer-annual-report-2014-womens-health

Davis, E. P., & Sandman, C. A. (2010). The timing of prenatal exposure to maternal cortisol and psychological stress is association with human infant cognitive development. *Child Development, 81,* 131-148. doi:10.1111/j.1467-8624.2009.01385.x

Davis, K., & Wardle, J. (1994). Body image and dieting in pregnancy. *Journal of Psychosomatic Research, 38,* 787-799. doi:10.1016/0022-3999(94)90067-1

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision, 1,* 79-101. doi:10.1037/dec0000004

Dawson, L., Rhodes, p., & Touyz, S. (2015). Defining recovery from anorexia nervosa: A Delphi study of expert practitioners' views. *Advances in Eating Disorders: Theory, Research, and Practice, 10,* 165-176. doi:10.1080/21662630.2015.1009145

Deeks, J. J., & Altman, D. G. (2004). Diagnostic tests 4: Likelihood ratios. *British Medical*

Journal, 329, 168-169. doi:10.1136/bmj.329.7458.168

DeJong, H., Fox, E., & Stein, A. (2016). Rumination and postnatal depression: A systematic review and a cognitive model. *Behaviour Research and Therapy, 82*, 38-49. doi:10.1016/j.brat.2016.05.003

Dellava, J. E., Von Holle, A., Torgersen, L., Reichborn-Kjennerud, T., Haugen, M., Meltzer, H. M., & Bulik, C. M. (2011). Dietary supplement use immediately before and during pregnancy in Norwegian women with eating disorders. *International Journal of Eating Disorders, 44*, 325-332. doi:10.1002/eat.20831

Delbecq, A. L., Van de Ven, A. H., & Gustafson, G. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott, Foresman and Company.

Deloitte. (2012). *The cost of perinatal depression in Australia*. Retrieved from: http://www.panda.org.au/images/stories/PDFs/PANDA_Exec_Summ_ Deloitte_Web.pdf

Department of Health and Ageing. (2013). *National Mental Health Report 2013: Tracking progress of mental health reform in Australia 1993 – 2011.* Commonwealth of Australia, Canberra.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed). Thousand Oaks, CA: Sage Publications.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting Delphi studies. *Journal of Clinical Epidemiology, 67*, 401-409. doi:10.1016/j.jclinepi.2013.12.002

Dickens, G. & Trethowan, W. H. (1971). Cravings and aversions during pregnancy. *Journal of Psychosomatic Research, 15*, 259-268 doi:10.1016/0022-3999(71)90037-7

DiPietro, J. A., Millet, S., Costigan, K. A., Gurewitsch, E., & Caulfield, L. E. (2003). Psychosocial influences on weight gain attitudes and behaviors during pregnancy. *Journal of the American Dietetics Association, 103,* 1314–1319. doi:10.1016/S0002-8223(03)01070-8

Donohoe, H. M., & Needham, R. D. (2008). Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. *International Journal of Tourism Research, 5,* 415-437. doi:10.1002/jtr.709

Downe, S., Finlayson, K., Tunçalp, Ö, & Gülmezoglu, A. M. (2015). What matters to women: A systematic scoping review to identify the processes and outcomes of antenatal care provision that are important to healthy pregnant women. *BJOG An International Journal of Obstetrics and Gynaecology, 123,* 529-539. doi:10.1111/1471-0528.13937

Duncombe, D., Wertheim, E. H., Skouteris, H., Paxton, S. J. & Kelly, L. (2008). How well do women adapt to changes in their body size and shape across the course of pregnancy? *Journal of Health Psychology, 13,* 503–515. doi:10.1177/1359105308088521

Easter, A. (2015). Understanding eating disorders in the antenatal and postnatal periods. *Perspective, 26,* 14-15.

Easter, A., Bye, A., Taborelli, E., Corfield, F., Schmidt, U., Treasure, J., & Micali, N. (2013). Recognising the symptoms: How common are eating disorders in pregnancy? *European Eating Disorders Review, 21,* 340-344. doi:10.1002/erc.2229

Easter, A., Solmi, F., Bye, A., Taborelli, E., Corfield, F., Schmidt, U., Treasure, J., & Micali, N. (2015). Antenatal and postnatal psychopathology among women with current and past eating disorders: Longitudinal patterns. *European Eating Disorders Review, 23,* 19-27. doi:10.1002/erv.2328

Easter, M. M. (2012). “Not all my fault”: Genetic, stigma, and personal responsibility for women with eating disorders. *Social Science & Medicine, 75,* 1408-1416. doi:10.1016/j.socscimed.2012.05.042

Eddy, K. T., Le Grange, D., Crosby, R. D., Hoste, R. R., Doyle, A. C., Smyth, A., & Herzog, D. B. (2011). Diagnostic classification of eating disorders in children and adolescents: How does DSM-IV-TR compare to empirically-derived categories? *Journal of the American Academy of Child and Adolescent Psychiatry, 49,* 277-293. doi:10.1016/j.jaac.2009.10.012

Edelstein, C. K., & King, B. H. (1992). Pregnancy and eating disorders. In J. Yager, & H. E. Gwirtsman (Eds.). *Special problems in managing eating disorders* (pp. 163-181). Washington, DC: American Psychiatric Association.

Elran-Barak, R., Zubery, E., & Steiner, E. (2014). *Can breast milk expression act as a compensation behavior in bulimia nervosa? Case report and literature review.* Presented at the International Conference on Eating Disorders, New York: United States.

Emery, R. L., Grace, J. L., Kolko, R. P., & Levine, M. S. (2017). Adapting the Eating Disorder Examination for use during pregnancy: Preliminary results from a community sample of women with overweight and obesity. *International Journal of Eating Disorders, 50,* 597-601. doi:10.1002/eat.22646

Eubank, B. H., Mohtadi, N. G., Lafave, M. R., Wiley, J. P., Bois, A. J., Boorman, R. S., & Sheps, D. M. (2016). Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Medical Research Methodology, 16,* 56. doi:10.1186/s12874-016-0165-8

Evans, J., Heron, J., Francomb, H., Oke, S., & Golding, J. (2001). Cohort study of depressed mood during pregnancy and after childbirth. *British Medical Journal, 323,* 257-260.

doi:10.1136/bmj.323.7307.257

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16,* 363–370. doi:10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-#

Fairburn, C. G., & Beglin, S. J. (2008). Eating Disorder Examination Questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 309–313). New York: Guilford Press.

Fairburn, C. G., Cooper, Z., Bohn, K., O'Connor, M. E., Doll, H. A., & Palmer, R. L. (2007). The severity and status of eating disorder NOS: Implications for DSM-V. *Behaviour Research and Therapy, 45,* 1705-1715. doi:10.1016/j.brat.2007.01.010

Fairburn, C. G., & Cooper, Z. (2011). Eating disorders, DSM-5 and clinical reality. *British Journal of Psychiatry, 198,* 8-10. doi:10.1192/bjp.bp.110.083881

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behavioural therapy for eating disorders: A transdiagnostic theory and treatment. *Behaviour Research and Therapy, 41,* 509-528. doi:10.1016/S0005-7967(02)00088-8

Fairburn, C. G., Stein, A., & Jones, R. (1992). Eating habits and eating disorders during pregnancy. *Psychosomatic Medicine, 54,* 665-672. doi:10.1097/00006842-199211000-00006

Fairburn, C. G., & Welch, S. L. (1990). The impact of pregnancy on eating habits and attitudes to shape and weight. *International Journal of Eating Disorders, 9,* 153–160. doi:10.1002/1098-108X(199003)9:2%3C153::AID-EAT2260090204%3E3.0.CO;2-8

Fairburn, C. G., & Welch, S. L., Doll, H. A., Davies, B. A., & O'Connor, M. E. (1997). Risk factors for bulimia nervosa. A community-based case-control study. *Archives of General Psychiatry, 54,* 509-517. doi:10.1001/archpsyc.1997.01830180015003

Farrar, D., Butterfield, G., Palethorpe, R., Jones, V., & Syson, J. (2013). Nutrition advice in pregnancy. *Practical Midwife, 16,* 13-17.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 105-146). New York: American Council on Education/Macmillan.

Fernandes, L., Hagen, K. B., Bijlsma, O., Andreassen, P., Christensen, P., Conaghan, P. G., . . . & Vliet Vlieland, T. P. M. (2012). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases, 72,* 1125-1135. doi:10.1136/annrheumdis-2012-202745

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine, 29,* 1043-1051. doi:10.1007/s00134-003-1761-8

Flynn, H. A., Henshaw, E., O'Mahen, H., & Forman, J. (2010). Patient perspectives on improving the depression referral processes in obstetrics settings: A qualitative study. *General Hospital Psychiatry, 32,* 9–16. doi:10.1016/j.genhosppsych.2009.07.005

Fornari, C., Dancyger, I., Renz, J., Skolnick, R., & Rochelson, B. (2014). Eating disorders and pregnancy: Proposed treatment guidelines for obstetricians and gynecologists. *Open Journal of Obstetrics and Gynecology, 4,* 90-94. doi:10.4236/ojog.2014.42016

Foster, S. F., Slade, P., & Wilson, K. (1996). Body image, maternal fetal attachment, and breastfeeding. *Journal of Psychosomatic Research, 41,* 181-184. doi:10.1016/0022-3999(96)00035-9

Fox, S., & Duggan, M. (2013). *Health online 2013.* Washington, DC: Pew Research Center's Internet & American Life Project.

Fox, P., & Yamaguchi, C. (1997). Body image change in pregnancy: A comparison of normal weight and overweight primigravidas. *Birth, 24,* 35–40. doi:10.1111/j.1523536X.

1997.tb00334.x

Franko, D. L., Blais, M. A., Becker, A. E., Delinsky, S. S., Greenwood, D. N., Flores, A. T., . . . & Herzog, D. B. (2001). Pregnancy complications and neonatal outcomes in women with eating disorders. *American Journal of Psychiatry 158,* 1461-1466 doi:10.1176/appi.ajp.158.9.1461

Franko, D. L, & Spurrell, E. B. (2000). Detection and management of eating disorders during pregnancy. *Obstetrics and Gynecology, 95,* 942–946. 10.1016/S0029-7844(00)00792-4

Franko, D. L., & Walton, B. E. (1993). Pregnancy and eating disorders: A review and clinical implications. *International Journal of Eating Disorders, 13,* 41-47. doi:10.1002/1098-108X(199301)13:1<41::AID-EAT2260130106>3.0.CO;2-L

Freizinger, M., Franko, D. L., Dacey, M., Okun, B., & Domar, A. D. (2010). The prevalence of eating disorders in infertile women. *Fertility and Sterility, 93,* 72-78. doi:10.1016/j.fertnstert.2008.09.055

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction.* Thousand Oaks, CA: Sage Publications.

Galletly, C., Castle, D., Dark, F., Humberstone, V., Jablensky, A., Killackey, E., . . . & Tran, N. (2016). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the management of schizophrenia and related disorders. *Australian and New Zealand Journal of Psychiatry, 50,* 410-472. doi:10.1177/0004867416641195

Garcia-Campayo, J., Sanz-Carrillo, C., Ibañez, J. A., Lou, S., Solano, V., & Alda, M. (2005). Validation of the Spanish version of the SCOFF questionnaire for the screening of eating disorders in primary care. *Journal of Psychosomatic Research, 59*, 51–55. doi:10.1016/j.jpsychores.2004.06.005

Garcia F. D., Grigioni S., Allais E., Houy-Durand E., Thibaut F., & Déchelotte P. (2011). Detection of eating disorders in patients: Validity and reliability of the French version of the SCOFF questionnaire. *Clinical Nutrition, 30*, 178–181. doi:10.1016/j.clnu.2010.09.007

Garcia F. D., Grigioni S., Chelali S., Meyrignac G., Thibaut F., & Dechelotte P. (2010). Validation of the French version of SCOFF questionnaire for screening of eating disorders among adults. *World Journal of Biological Psychiatry, 11*, 888–893. doi:10.3109/15622975. 2010.483251

Garner, D. M. (1991). *Eating Disorder Inventory-2: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Garner, D. M. (2004). *Eating Disorder Inventory-3: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Gavin, N.I., Gaynes, B.N., Lohr, K.N., Meltzer-Brody, S., Gartlehner, G., & Swinson, T. (2005). Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics and Gynecology, 106*, 1071-1083. doi:10.1097/01.AOG.0000183597.31630.db

Gemmill, A. W., Leigh, B., Ericksen, J., & Milgrom, J. (2006). A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health, 6*, 211. doi:10.1186/1471-2458-6-211

Giardinelli, L., Cecchelli, C., & Innocenti, A. (2008). Psychiatric disorders in pregnancy. *Journal of Psychopathology, 14*, 211-219.

Gipson, J. D., Koenig, M. A., Hindin, M. J. (2008). The effects of unintended pregnancy on infant, child, and parental health: A review of the literature. *Studies in Family Planning, 39,* 18-38. doi:10.1111/j.1728-4465.2008.00148.x

Golding, J., Pembrey, M., & Jones, R. (2001). ALSPAC – the Avon longitudinal study of parents and children: Study methodology. *Paediatric and Perinatal Epidemiology, 15,* 74-87. doi:10.1046/j.1365-3016.2001.00325.x

Gonçalves, S., Freitas, F., Freitas-Rosa, M. A., & Machado, B. C. (2015). Dysfunctional eating behaviour, psychological well-being and adaptation to pregnancy: A study with women in the third trimester of pregnancy. *Journal of Health Psychology, 20,* 535-542. doi:10.1177/1359105315573432

Goodman, C. M. (1987). The Delphi technique: A critique. *Journal of Advanced Nursing, 12,* 729–734. doi:10.1111/j.1365-2648.1987.tb01376.x

Goodwin, A., Astbury, J., & McMeeken, J. (2000). Body image and psychological wellbeing in pregnancy. A comparison of exercisers and non-exercisers. Australian and *New Zealand Journal of Obstetrics and Gynaecology, 40,* 442-447. doi:10.1111/j.1479-828X.2000.tb01178.x

Gordon, K. H., Perez, M., & Joiner, T. E. (2002). The impact of racial stereotypes on eating disorder recognition. *International Journal of Eating Disorders, 32,* 219–224. doi:10.1002/eat.10070

Götestam, K. G., & Agras, W. S. (1995). General population-based epidemiological study of eating disorders in Norway. *International Journal of Eating Disorders, 18,* 119–126. doi:10.1002/1098-108X(199509)18:2%3C119::AID-EAT2260180203%3E3.0.CO;2

Gough, K. (1998). *Eating habits and body image during pregnancy* (doctoral dissertation). Retrieved from: ProQuest Dissertations and Theses database (UMI No. 10011131).

Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting, 27,* 183-195. doi:10.1016/j.ijforecast.2010.05.004

Graham, B., Regehr, G., & Wright, J. G. (2003). Delphi as a method to establish consensus for diagnostic criteria. *Journal of Clinical Epidemiology, 56,* 1150-1156. doi:10.1016/S0895-4356(03)00211-7

Gray, R. (2013). *Life stage: Pre-conception and pregnancy. Annual report of the chief medical officer 2012. Our children deserve better: Prevention pays*. London: Department of Health. Retrieved from: http://socialwelfare.bl.uk/subjectareas/services-client-groups/children-youngpeople/departmentofhealth/15526533571_2901304_CMO_All.pdf

Greatorex, J., & Dexter, T. (2000). An accessible analytical approach for investigating what happens between the rounds of a Delphi study. *Journal of Advanced Nursing, 32,* 1016–1024. doi:10.1046/j.1365-2648.2000.01569.x

Green B., Jones M., Hughes D. & Willimas A. (1999) Applying the Delphi technique in a study of GP's information requirements. *Health and Social Care in the Community, 7,* 198-205. doi:10.1046/j.1365-2524.1999.00176.x

Green, S.B., Lissitz, R.W., & Mulaik, S.A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-838. doi:10.1177/001316447703700403

Greenhalgh, T. (1997). How to read a paper: Papers that report diagnostic or screening tests. *British Medical Journal, 315,* 540. doi:10.1136/bmj.315.7107.540

Greenhalgh, T., Howick., J., & Maskrey, N. (2014). Evidence based medicine: A movement in crisis? *British Medical Journal, 348,* 3725. doi:10.1136/bmj.g3725

Gregory, R. J. (2015). *Psychological testing: History, principles, and applications* (7th ed.). Harlow, Essex: Pearson Education Limited.

Grilo, C. M., Masheb, R. M., & Wilson, G. T. (2001). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder.

Journal of Consulting and Clinical Psychology, 69, 317–322. doi:10.1037/0022-006X.69.2.317

Grilo, C. M., Pagano, M. E., Skodol, A. E., Sanislow, C. A., McGlashan, T. H., Gunderson, J. G., & Stout, R. L. (2007). Natural course of bulimia nervosa and of eating disorder not otherwise specified: 5-year prospective study of remissions, relapses, and the effects of personality disorder psychopathology. *Journal of Clinical Psychiatry, 68,* 738-746. doi:10.4088/JCP.v68n0511

Gustafson, D. H., Shukla, R. K., Delbecq, A., & Walstre, G. W. (1973). A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organisational Behaviour and Human Performance, 9,* 280–291. doi:10.1016/0030-5073(73)90052-4

Haedt, A., & Keel, P. (2007). Maternal attachment, depression, and body dissatisfaction in pregnant women. *Journal of Reproductive and Infant Psychology, 25*, 285-295. doi:10.1080/02646830701691327

Hajian-Tilkai, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*, 627-635.

Hall-Flavin, D. K. (2015). *Is pregorexia for real?* Retrieved from: http://www.mayoclinic.org/healthy-lifestyle/pregnancy-week-by-week/expert-answers/pregorexia/faq-20058356

Hansson, E., Daukantaité, D., & Johnsson, P. (2015). SCOFF in a general Swedish adolescent population. *Journal of Eating Disorders, 3,* 48-56. doi:10.1186/s40337-015-0087-6

Hardy, D. J., O'Brien, A. P., Gaskin, C. J., O'Brien, A. J., Morrison-Ngatal, E., Skews, G., Ryan, T., & McNulty, N. (2004). Practical application of the Delphi technique in a

bicultural mental health nursing study in New Zealand. *Journal of Advanced Nursing, 46*, 95–109. doi:10.1111/j.1365-2648.2003.02969.x

Harris, A. A. (2010). Prenatal advice for caring for women with eating disorders during the perinatal period. *Journal of Midwifery and Womens Health, 55,* 579-586. doi:10.1016/j.mwh.2010.07.008

Hart, L. M., Damiano, S. R., Chittleborough, P., Paxton, S., & Jorm, A. F. (2014). Parenting to prevent body dissatisfaction and unhealthy eating patterns in preschool children: A Delphi consensus study. *Body Image, 11*, 418-425. doi:10.1016/j.bodyim.2014.06.010

Harvey, A.G., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment.* Oxford: Oxford University Press.

Hasson, F., & Keeney, S. (2011). Enhancing the rigour in the Delphi technique research. *Technological Forecasting and Social Change, 78,* 1695-1704. doi:10.1016/j. techfore.2011.04.005

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*, 1008–1015. doi:10.1046/j.1365-2648.2000.01567.x

Hautala, L., Junnila, J., Alin, J., Gronroos, M., Maunula, A. M., Karukivi, M., Liuksila, P. R., . . . & Saarjiävi, S. (2009). Uncovering hidden eating disorders using the SCOFF questionnaire: Cross-sectional survey of adolescent and comparisons with nurse assessments. *International Journal of Nursing Studies, 46,* 1439-1447. doi:10.1016/j.ijnurstu.2009.04.007

Hawkins, L. K., & Gottlieb, B. (2013). Screening for eating disorders in pregnancy: How uniform screening during a high risk period could minimise under-recognition. *Journal of Women's Health, 22,* 390-392. doi:10.1089/jwh.2013.4313

Hay, P., Buttner, P., Mond, J., Paxton, S. J., Rodgers, B., Quirk, F., & Darby, A. (2010). Quality of life, course and predictors of outcomes in community women with EDNOS and common eating disorders. *European Eating Disorders Review, 18,* 281-295. doi:10.1002/erv.1023

Hempel, I., Buck, N., Cima, M., & van Marle, H. (2013). Review of risk assessment instruments in juvenile sex offenders: What is next? *International Journal of Offender and Comparative Criminology, 57*, 208-228. doi:10.1177/0306624X11428315

Heron, J., O'Connor, T. G., Evans, J., Golding, J., Glover, V., & ALSPAC Study Team. (2004). The course of anxiety and depression through pregnancy and the postpartum in a community sample. *Journal of Affective Disorders, 80*, 65-73. doi:10.1016/j.jad. 2003.08.004

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions (version 5.1.0). The Cochrane collaboration*. Retrieved from: www.cochrane-handbook.org.

Highet, N. J., Gemmil, A., & Milgrom, J. (2010). Depression in the perinatal period: Awareness, attitudes and knowledge in the Australian population. *Australian New Zealand Journal of Psychiatry, 45*, 223-231. doi:10.3109/00048674.2010.547842

Highet, N. J., & Purtell, C. A. (2012). *The National Perinatal Depression Initiative: A synopsis of progress to date and recommendations for beyond 2013.* Melbourne, VIC: beyondblue.

Highet, N. J., Stevenson, A., Purtell, C. A., & Coo Calcagni, S. (2014). Qualitative insights into women's personal experiences of perinatal depression and anxiety. *Women Birth, 27*, 179-184. doi:10.1016/j.wombi.2014.05.003.

Hilbert, A., de Zwaan, M., & Braehler, E. (2012). How frequent are eating disturbances in the population? Norms of the Eating Disorder Examination-Questionnaire. *PLoS One, 7*, e29125. doi:10.1371/journal.pone.0029125

Hill L. S., Reid F., Morgan J. F., & Lacey J. H. (2010). SCOFF, the development of an eating disorder screening questionnaire. *International Journal of Eating Disorders, 43*, 344–351. doi:10.1002/eat.20679

Hogg, S. (2013). *Prevention in mind: all babies count: Spotlight on perinatal mental illness*. London: NSPCC. Retrieved from: https://www.nspcc.org.uk/services-and-resources/research-and-resources/2013/all-babies-count-spotlight-perinatal-mental-health/

Hogue, M-C. B., Doran, E., & Henry, D. A. (2012). A prompt to the web: The media and health information seeking behaviour. *PLoS ONE, 7*, e34314. doi:10.1371/journal.pone.0034314

Holey, E. A., Feeley, J. L., Dixon, J., & Whittaker, V. J. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology, 7*. doi:10.1186/1471-2288-7-52.

Hollifield, J., & Hobdy, J. (1990). The course of pregnancy complicated by bulimia. *Psychotherapy, 27,* 249-255. doi:10.1037/0033-3204.27.2.249

Hooper, D. E. (1996). Detecting GD and preeclampsia. Effectiveness of routine urine screening for glucose and protein. *Journal of Reproductive Medicine, 41*, 885-888.

Hou, R., Mogg, K., Bradley, B., Moss-Morris, R., Peveler, R., & Roefs, A. (2011). External eating, impulsivity, and attentional bias to food cues. *Appetite, 56,* 424-427. doi:10.1016/j.appet.2011.01.019

Howard, L. M., Goss, C., Leese, M., & Thornicroft, G. (2003). Medical outcome of pregnancy in women with psychotic disorders and their infants in the first year after birth. *British Journal of Psychiatry, 182,* 63–67. doi:10.1192/bjp.182.1.63

Howard, L. M., Molyneaux, E., Dennis, C-L., Rochat, T., Stein, A., & Milgrom, J. (2014). Non-psychotic mental disorders in the perinatal period. *Lancet, 384,* 1775–1788. doi:10.1016/S0140-6736(14)61276-9

Hrabosky, J. I., White, M. A., Masheb, R. M., Rothschild, B. S., Burke-Martindale, C. H., & Grilo, C. M. (2008). Psychometric evaluation of the Eating Disorder Examination-Questionnaire for bariatric surgery candidates. *Obesity, 16,* 763–769. doi:10.1038/oby.2008.3

Hsu, L. G. (1989). The gender gap in eating disorders: Why are the eating disorders more common among women? *Clinical Psychological Review, 9,* 393–407. doi:10.1016/0272-7358(89)90063-9

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation, 12,* 1-7. doi:10.4018/978-1-4666-0074-4.ch011

Hubin-Gayte, M., & Squires, C. (2012). Étude de l'impact de la grossesse sur les comportements alimentaires à travers l'utilisation du questionnaire SCOFF. *L'Evolution Psychiatrique, 77,* 201-212. doi:10.1016/j.evopsy.2012.01.002

Hudson, J., Hiripi, E., Pope Jr., H., & Kessler, R. (2007). The prevalence and correlates of eating disorders in the national comorbidity survey replication. *Biological Psychiatry, 61,* 348-358. doi:10.1016/j.biopsych.2006.03.040

Institute of Medicine. (2010). *Weight gain during pregnancy: Re-examining the guid*elines. Washington, DC: National Academy Press.

Iqbal, S., & Pipon-Young, L. (2009). The Delphi method. *Methods, 22,* 598-601.

Jacobi, C., Abascal, L., & Taylor, C. (2004). Screening for eating disorders and high-risk behavior: Caution. *International Journal of Eating Disorders, 36,* 280-295. doi:10.1002/eat.20048

Joanna Briggs Institute., & University of Adelaide. (2013). *New JBI levels of evidence*. Retrieved from: http://joannabriggs.org/assets/docs/approach/JBI-Levels-of-evidence_2014.pdf

Jokinen, J., Nordström, A. L., & Nordström, P. (2008). ROC analysis of dexamethasone suppression test threshold in suicide prediction after attempted suicide. *Journal of Affective Disorders, 106,* 145–152. doi:10.1016/j.jad.2007.06.008

Jones, W. P., & Loe, S. A. (2013). Optimal number of questionnaire response categories: more may not be better. *Sage Open, 3,* 1–10. doi:10.1177/2158244013489691

Jönsson, M. L., Hahn, U., & Olsson, E. J. (2015). The kind of group you want to belong to: Effects of group structure on group accuracy. *Cognition, 142,* 191-204. doi:10.1016/j.cognition.2015.04.013

Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry, 49,* 887-897. doi:10.1177/0004867415600891

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 32-36. doi:10.1007/BF02291575

Kapadia, M. Z., Gaston, A., Van Blyderveen, S., Schmidt, L., Beyene, J., McDonald H., & McDonald, S. D. (2015). Psychological antecedents of excess gestational weight gain: A systematic review. *BMC Pregnancy Childbirth, 15,* 107. doi: 10.1186/s12884-015-0535-y

Keel, P. K., Brown, T. A., Holm-Denoma, J., & Bodell, L. P. (2011). Comparison of DSM-IV versus proposed DSM-5 diagnostic criteria for eating disorders: reduction of eating

disorder not otherwise specified and validity. *International Journal of Eating Disorders, 44,* 553-560. doi:10.1002/eat.20892

Keenan, K., Bartlett, T. Q., Nijland, M., Rodriguez, J. S., Nathanielsz, P. W., & Zürcher, N. R. (2013). Poor nutrition during pregnancy and lactation negatively affects neurodevelopment of the offspring: Evidence from a translational primate model. *American Journal of Clinical Nutrition, 98*, 396–402. doi:10.3945/ajcn.112.040352

Keeney, S., Hasson, F., & McKenna, H. (2001). A critical review of the Delphi technique as a research methodology for nursing. *International Journal of Nursing Studies, 38*, 195-200. doi:10.1016/S0020-7489(00)00044-4

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing, 53*, 205-212. doi:10.1111/j.1365-2648.2006.03716.x

Keller, M. B., Lavori, P. W., Freidman, B., Neilson, E., Endicott, J., McDonald-Scott, P., & Andearson, N. C. (1987). The longitudinal interval follow-up evaluation. *Archives of General Psychiatry, 44*, 540-548. doi:10.1001/archpsyc.1987.01800180050009

Kelly, C. M., Jorm, A. F., Kitchener, B. A., & Langlands, R. L. (2008). Development of mental health first aid guidelines for suicidal ideation and behaviour: A Delphi study. *BioMed Central Psychiatry, 8*. doi:10.1186/1471-244X-8-17

Kendell, R. E., Wainwright, S., Hailey, A., & Shannon, B. (1976). The influence of childbirth on psychiatric morbidity. *Psychological Medicine, 6,* 297-302. doi:10.1017/S003329 17000138

Kennedy, H. (2004). Enhancing Delphi research: Methods and results. *Journal of Advanced Nursing, 45*, 504-511. doi:10.1046/j.1365-2648.2003.02933.x

Kills Small, N. J., Simons, J. S., & Stricherz, M. (2007). Assessing criterion validity of the Simple Screening Instrument for Alcohol and Other Drug Abuse (SSI-AOD) in a college population. *Addictive Behaviors, 32,* 2425–2431. doi:10.1016/j.addbeh.2007.04.003

King, J. C. (2000). Physiology of pregnancy and nutrient metabolism. *American Journal of Clinical Nutrition, 71,* 1218-1225.

Kingston, D., Austin, M-P., McDonald, S. W., Vermeyden, L., Heaman, M., Hegadoren, K., Lasiuk, G., . . . & Biringer, A. (2015). Pregnant women's perceptions of harms and benefits of mental health screening. *PLoS ONE, 10,* 145-189. doi:10.1371/journal.pone.0145189

Kjelsäs, E., Bjørnstrøm, C., & Götestam, K. G. (2004). Prevalence of eating disorders in female and male adolescents (14-15 years). *Eating Behaviours, 5,* 13-25. doi:10.1016/S1471-0153(03)00057-6

Klemchuk, H. P., Hutchinson, C. B., & Frank, R. I. (1990). Body dissatisfaction and eating-related problems on the college campus: Usefulness of the Eating Disorder Inventory with a nonclinical population. *Journal of Counseling Psychology, 37,* 297-305. doi:10.1037/0022-0167.37.3.297

Klima, C. S. (1998). Unintended pregnancy: Consequences and solutions for a worldwide problem. *Journal of Nurse-Midwifery, 43,* 483-491. doi:10.1016/S0091-2182(98)00063-9

Kline, T. J. B. (2005). *Psychological testing: A practical to design and evaluation*. Thousand Oaks, CA: Sage Publications.

Klump, K. L., Bulik, C. M., Kaye, W. H., Treasure, J., & Tyson, E. (2009). Academy for eating disorders position paper: Eating disorders are serious mental illnesses. *International Journal of Eating Disorders, 42,* 97-103. doi:10.1002/eat.20589

Knoph Berg, C., Torgersen, L., Von Holle, A., Hamer, R. M., Bulik, C. M., & Reichborn Kjennerud, T. (2011). Factors associated with binge eating disorder in pregnancy. *International Journal of Eating Disorders, 44,* 124-133. doi:10.1002/eat.20797

Knoph, C., Von Holle, A., Zerwas, S., Torgersen, L., Tambs, K., Stoltenberg, C., Bulik, C. M., & Reichborn-Kjennerud, T. (2013). Course and predictors of maternal eating disorders in the postpartum period. *International Journal of Eating Disorders, 46,* 355-368. doi:10.1002/eat.22088

Kolko, R. P., Emery, R. L., Marcus, M. D., & Levine, M. D. (2017). Loss of control over eating before and during early pregnancy among community women with overweight and obesity. *International Journal of Eating Disorders, 50,* 582-586. doi:10.1002/eat.22630

Koubaa, S., Hallstrom, T., Lindholm, C., & Hirschberg ,A. L. (2005). Pregnancy and neonatal outcomes in women with eating disorders. *Obstetrics and Gynecology, 105,* 255–260. doi:10.1097/01.AOG.0000148265.90984.c3

Koupil, I., Tooth, L., Heshmati, A., & Mishra, G. (2016). Social patterning of overeating, binge eating, compensatory behaviours and symptoms of bulimia nervosa in young adult women: Results from the Australian Longitudinal Study on Women's Health. *Public Health Nutrition, 19,* 3158-3168. doi:10.1017/S1368980016001440

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2,* 151-160. doi:10.1007/BF02288391

Lacey, J. H., & Smith, G. (1987). Bulimia nervosa: The impact of pregnancy on mother and baby. *British Journal of Psychiatry, 150,* 777. doi:10.1192/bjp.150.6.777

Lähteenmäki S., Aalto-Setälä T., Suokas J. T, Saarni S. E., Perälä J., Saarni S. I., . . . & Suvisaari J. M. (2009). Validation of the Finnish version of the SCOFF questionnaire

among young adults aged 20 to 35 years. *BMC Psychiatry*, *9*. doi:10.1186/1471-244X-9-5

Lai, L., Flower, Moore, M., & Lewith, G. (2015). Developing clinical practice guidelines for Chinese herbal treatment of polycystic ovary syndrome: A mixed-methods modified Delphi study. *Complementary Therapies & Medicine, 23*, 430-438. doi:10.1016/j.ctim.2015.03.003

Lai, B. P.-Y., Tang, C. S.-K., & Tse, W. K.-L. (2006). A longitudinal study investigating disordered eating during the transition to motherhood among Chinese women in Hong Kong. *International Journal of Eating Disorders, 39*, 303-311. doi:10.1002/eat.20266

Lai, B. P.-Y., Tang, C. S.-K., & Tse, W. K.-L. (2005). Prevalence and psychosocial correlates of disordered eating among Chinese pregnant women in Hong Kong. *Eating Disorders, 13*, 171-186. doi:10.1080/10640260590918991

Lande, B., Andersen, L. F., Bærug, A., Trygg, K. U., Lund-Larsen, K., Veierød, M. B., & Aa Bjørneboe, G. E. (2003). Infant feeding practices and associated factors in the first six months of life: The Norwegian Infant Nutrition Survey. *Acta Pædiatricia, 92*, 152-161. doi:10.1111/j.1651-2227.2003.tb00519.x

Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008). First aid for depression: A Delphi consensus study with consumers, carers and clinicians. *Journal of Affective Disorders, 105*, 157–166. doi:10.1016/j.jad.2007.05.004

Larsson, G., & Andersson-Ellström, A. (2003). Experiences of pregnancy-related body shape changes and of breastfeeding in women with a history of eating disorders. *European Eating Disorders Review, 11*, 116-124. doi:10.1002/erv.497

Laurent, H. K., Leve, L. D., Neiderhiser, J. M., Natsuaki, M. B., Shaw, D. S., Harold, G. T., & Reiss, D. (2013). Effects of prenatal and postnatal parent depressive symptoms on

adopted child HPA regulation: independent and moderated influences. *Developmental Psychology, 49*, 876-886. doi:10.1037/a0028800

Lawson, K., & Tulloch, M. I. (1995). Breastfeeding duration: Prenatal intentions and postnatal practices. *Journal of Advanced Nursing, 22*, 841-849. doi:10.1111/j.1365-2648.1995.tb02633.x

Leddy, M. A., Jones, C., Morgan, M. A., & Schulkin, J. (2009). Eating disorders and obstetric-gynecologic care. *Journal of Women's Health, 18,* 1395-1401. doi:10.1089/jwh.2008.1183

Leigh, B., & Milgrom, J. (2007). Acceptability of antenatal screening for depression in routine antenatal care. *Australian Journal of Advanced Nursing, 4*, 14–18.

Lemberg, R., & Phillips, J. (1989). The impact of pregnancy on anorexia nervosa and bulimia. *International Journal of Eating Disorders, 8*, 285–295. doi:10.1002/1098-108X(198905)8:3%3C285::AID-EAT2260080304%3E3.0.CO;2-P10.1080/02646830903190904

Leung, S. F., Lee, K. L., Lee, S. M., Leung, S. C., Hung, W. S., Lee, W. L., Leung, , Y. Y., . . . & Wong, Y. N. (2009). Psychometric properties of the SCOFF questionnaire (Chinese version) for screening eating disorders in Hong Kong secondary school students: A cross-sectional study. *International Journal of Nursing Studies, 46,* 239-247. doi:10.1016/j.ijnurstu.2008.09.004

Leung, S. S. L., Leung, C., Lam, T. H., Hung, S. F., Chan, R., Yeung, T., Miao, M., . . . & Lee, D. T. S. (2011). Outcome of a postnatal depression screening programme using the Edinburgh Postnatal Depression Scale: A randomized controlled trial. *Journal of Public Health, 33*, 292-301. doi:10.1093/pubmed/fdq075

Linacre, J. M. (2014). *Facets – Rasch measurement computer program*. Chicago, IL: MESA Press.

Linna, M. S., Raevuori, A., Haukka, J., Suvisaari, J. M., Suokas, J. T., & Gissler, M. (2014). Reproductive health outcomes in eating disorder. *International Journal of Eating Disorders, 46,* 826-833. doi:10.1002/eat.22179

Linstone, H. A., & Turoff, M. (1975). Introduction. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 3-12). Reading, MA: Addison-Wesley Publishing Company.

Lisonkova, S., Muraca, G. M., Potts, J., Liauw, J., Chan, W-S., Skoll, A., & Lim, K. I. (2017). Association between pre-pregnancy body mass index and severe maternal morbidity. *Journal of the American Medical Association, 318,* 1777-1786. doi:10.1001/jama.2017.16191

Loewenthal, K. M. (1996). *An introduction to psychological tests and scales.* London: UCL Press.

Lorenz, J., Rauhut, H., Schweitzer F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences, 108*, 9020–9025.

Lovestone, S., & Kumar, R., (1993). Postnatal psychiatric illness: The impact on partners. *British Journal of Psychiatry, 163*, 210-216. doi:10.1192/bjp.163.2.210

Lowes, H., Kopeika, J., Micali, N., & Ash, A. (2012). Anorexia nervosa in pregnancy. *The Obstetrician and Gynaecologist, 14,* 179-187. doi:10.1111/j.1744-4667.2012.00110.x

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4,* 73–79. doi:10.1027/1614-2241.4.2.73

Luce, K. H., Crowther, J. H., & Pole, M. (2008). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders, 41*, 273–276. doi:10.1002/eat.20504

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi Methodology? *Journal of Extension, 35,* 10-11.

Lydsdottir, L. B., Howard, L. M., Olafsdottir, H., Thome, M., Tyrfingsson, P., & Sigurdsson, J. F. (2014). The mental health characteristics of pregnant women with depressive symptoms identified by the Edinburgh Postnatal Depression Scale. *Journal of Clinical Psychiatry, 75,* 393–398. doi:10.4088/JCP.13m08646

Machado, P. P., Gonçalves, S., & Hoek, H. W. (2013). DSM-5 reduces the proportion of EDNOS cases: evidence from community samples. *International Journal of Eating Disorders, 46,* 60-65. doi:10.1002/eat.22040

Machado, P. P., Machado, B. C., Gonçalves, S., & Hoek, H. W. (2007). The prevalence of eating disorders not otherwise specified. *International Journal of Eating Disorders, 40,* 212-217. doi:10.1002/eat.20358

Machado, P. P., Martins, C., Vaz, A. R., Conceicao, E., Bastos, A. P., & Goncalves, S. (2014). Eating Disorder Examination Questionnaire: Psychometric properties and norms for the Portuguese population. *European Eating Disorders Review, 22,* 448–453. doi:10.1002/erv.2318.

Magnus, P., Irgens, L. M., Haug, K., Nystad, W., Skjaerven, R., Stoltenberg, C., & MoBa Study Group. (2006). Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology, 35,* 1146-1150. doi:10.1093/ije/dyl170

Manikandan, S. (2011). Measures of central tendency: Median and mode. *Journal of Pharmacology and Pharmacotherapy, 2,* 214-215. doi:10.4103/0976-500X.83300

Marquer, C., Barry, C., Mouchenik, Y., Hustache, S., Djibo, D., & Manzo, M. (2012). A rapid screening tool for psychological distress in children 3–6 years old: Results of a validation study. *BMC Psychiatry, 12,* 170. doi:10.1186/1471-244x-12-170

Marques, L., Alegria, M., Becker, A. E., Chen, C. N., Fang, A., Chosak, A., & Diniz, J. B. (2011). Comparative prevalence, correlates of impairment, and service utilization for eating disorders across US ethnic groups: Implications for reducing ethnic disparities in health care access for eating disorders. *International Journal of Eating Disorders, 44*, 412-420. doi:10.1002/eat.20787

Mason, Z., Cooper, M., & Turner, H. (2012). The experience of pregnancy in women with a history of anorexia nervosa: An interpretive phenomenological analysis. *Journal of Behavioural Addictions, 1,* 59-67. doi:10.1556/JBA.1.2012.2.3

Mathieu, J. (2009). What is pregorexia? *Journal of the American Dietetic Association, 21,* 976-979. doi:10.1016/j.jada.2009.04.021

Matthey, S., White, T., Phillips, J., Taouk, R., Chee, T. T., & Barnett, B. (2005). Acceptability of routine antenatal psychosocial assessments to women from English and non-English speaking backgrounds. *Archives of Women's Mental Health, 8*, 171–180. doi:10.1007/s00737-005-0076-x

Mayo, K., Melamed, N., Vandenberghe, H., & Berger, H. (2015). The impact of adoption of the international association of diabetes in pregnancy study group criteria for the screening and diagnosis of gestational diabetes. *American Journal of Obstetrics and Gynecology, 212*, 221-229. doi:10.1016/j.ajog.2014.08.027

Mazzeo, S. E., Slof-Op't Landt, M. C., Jones, I., Mitchell, K., Kendler, K. S., Neale, M. C., Aggen, S. H., & Bulik, C. M. (2006). Associations among postpartum depression, eating disorders, and perfectionism in a population-based sample of adult women. *International Journal of Eating Disorders*, *39*, 202-211. doi:10.1002/eat.20243

McFarlane, L., Owens, G., & Del Pozo Cruz, B. (2016). Identifying the features of an exercise addiction: A Delphi study. *Journal of Behavioral Addictions, 5*, 474-484. doi:10.1556./2006.5.2016.060

McGee, S. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine, 17*, 647-650. doi:10.1046/j.1525-1497.2002.10750.x

McKenna, H. (1994). The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing, 19*, 1221-1225. doi:10.1111/j.1365-2648.1994.tb01207.x

McKenna, H., & Keeney, S. (2008). Delphi studies. In R. Watson, H. P. McKenna, S. Cowman, & J. Keady (Eds.), *Nursing research: Designs and* methods (pp. 251-260). Edinburgh: Churchill Livingstone.

Mead, D., & Moseley, L. (2001). The use of Delphi as a research approach. *Nurse Researcher, 8*, 4-23. doi:10.7748/nr2001.07.8.4.4.c6162

Melander, H. L., & Lauri, S. (1999). Fears associated with pregnancy and childbirth: Experiences of women who have recently given birth. *Midwifery, 15*, 177-182. doi:10.1016/S0266-6138(99)90062-1

Meltzer-Brody, S., & Stuebe, A. (2014). The long-term psychiatric and medical prognosis of perinatal mental illness. *Best Practice & Research: Clinical Obstetrics & Gynaecology, 28*, 49–60. doi:10.1016/j.bpobgyn.2013.08.009

Meltzer-Brody, S., Zerwas, S., Leserman, J., Von Holle, A., Regis, T., & Bulik, C. (2011). Eating disorders and trauma history in women with perinatal depression. *Journal of Women's Health, 20,* 863. doi:10.1089/jwh.2010.2360

Micali, N. (2010). Management of eating disorders during pregnancy. *Progress in Neurology and Psychiatry, 23*, 635–641. doi:10.1002/pnp.158

Micali, N., Simonoff, E., & Treasure, J. (2009). Infant feeding and weight in the first year of life in babies of women with eating disorders. *Journal of Pediatrics, 154*, 55-60. doi:10.1016/j.jpeds.2008.07.003

Micali, N., Simonoff, E., & Treasure, J. (2007). Risk of major adverse perinatal outcomes in women with eating disorders. *British Journal of Psychiatry, 190,* 255–259. doi:10.1192/bjp.bp.106.020768

Micali, N., Treasure, J., & Simonoff, E. (2007). Eating disorders symptoms in pregnancy: A longitudinal study of women with recent and past eating disorders and obesity. *Journal of Psychosomatic Research, 63,* 297–303. doi:10.1016/j.jpsychores.2007.05.003

Micali, N. & Treasure, J. (2009). Biological effects of maternal ED on pregnancy and foetal development: A review. *European Eating Disorders Review, 17,* 448–454. doi:10.1002/erv.963

Miller, S., Abaolos, E. Chamillard, M., Ciapponi, A., Colaci, D., Comandé, D., . . . & Althabe, F. (2016). Beyond too little, too late and too much, too soon: A pathway towards evidence-based, respectful maternity care worldwide. *Lancet, 388,* 2176-2192. doi:10.1016/S0140-6736(16)31472-6

Miller, L. A., Lovler, R. L., & McIntire, S. A. (2012). *Foundations of psychological testing: A practical approach* (4th ed.). Thousand Oaks, CA: Sage Publications.

Mintz, L. B., & O'Halloran, M. S. (2000). The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment, 74,* 489-503. doi:10.1207/S15327752JPA7403_11

Mitchell, A. M., & Bulik, C. M. (2006). Eating disorders and women's health: An update. *Journal of Midwifery and Women's Health, 51,* 193-201. doi:10.1016/j.jmwh.2006.01.005

Mitchell, A. J., & Coyne, J. (2009). Screening for postnatal depression: barriers to success. *BJOG: An International Journal of Obstetrics and Gynaecology, 116,* 11-14. doi:10.1111/j.1471-0528.2008.01834.x

Mitchell, J. E., & Crow, S. (2006). Medical complications of anorexia nervosa and bulimia nervosa. *Current Opinion in Psychiatry, 19,* 438-443. doi:10.1097/01.yco.0000228768.79097.3e

Mitchell, P. B., Johnston, A. K., Frankland, A., Slade, T., Green, M. J., Roberts, G., . . . & Hadzi-Pavlovic, D. (2013). Bipolar disorder in a national survey using the World Mental Health Version of the Composite International Diagnostic Interview: The impact of differing diagnostic algorithms. *Acta Psychiatrica Scandinavica, 127,* 381-393. doi:10.1111/acps.12005

Mitroff, I. I., & Turoff, M. (1975). Philosophical and methodological foundations of Delphi. In H. A. Linstone & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 17-35). Reading, MA: Addison-Wesley Publishing Co.

Mitsui, T., Toshida, T., & Komaki, G. (2017). Psychometric properties of the Eating Disorder Examination-Questionnaire in Japanese adolescents. *BioPsychoSocial Medicine, 11,* 1-9. doi:10.1186/s13030-017-0094-8

Mittnacht, A. M., & Bulik, C. M. (2015). Best nutrition counseling practices for the treatment of anorexia nervosa: A Delphi study. *International Journal of Eating Disorders, 48,* 111-122. doi:10.1002/eat.22319

Moalla, M., Feki, I., Boudabbous, J., Masmoudi, J., & Jaoua, A. (2015). Impact of pregnancy on eating behaviours. *European Psychiatry, 30,* 38-31. doi:10.1016/S0924-9338(15)31037-3

Mohamadirizi, S., Kordi, M., Shakeri, M. T., & Modares-Gharavi, M. (2015). The relationship between eating disorder symptoms and obsessive compulsive disorder in primigravida women. *Iranian Journal of Nursing and Midwifery Research, 20,* 642-646. doi:10.4103/1735- 9066.170015

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62,* 1006-1012. doi:10.1371/journal.pmed.1000097

Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Temporal stability of the Eating Disorder Examination Questionnaire. *International Journal of Eating Disorders, 36,* 195-203. doi:10.1002/eat.20017

Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy, 42,* 551–567. doi:10.1016/S0005-7967(03)00161-X

Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy, 44,* 53–62. doi:10.1016/j.brat.2004.12.003

Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Self-recognition of disordered eating among women with bulimic-type eating disorders: A community-based study. *International Journal of Eating Disorders*, *39*, 747–753. doi:10.1002/eat.20306

Mond, J. M., Myers, T. C., Crosby, R. D., Hay, P. J., Rodgers, B., Morgan, J. F., Lacey, J. H., & Mitchell, J. E. (2008). Screening for eating disorders in primary care: EDE-Q versus SCOFF. *Behaviour Research and Therapy*, *46*, 612–622. doi:10.1016/j.brat. 2008.02.003

Morgan, J. F. (1999). Eating disorders and reproduction. *Australia New Zealand Obstetrics and Gynecology, 39,* 167-173. doi:10.1111/j.1479-828X.1999.tb03364.x

Morgan, J. F. (1999). Eating disorders and gynecology: Knowledge and attitudes among clinicians. *Acta Obstetricia et Gynecologica Scandinavica*, *78*, 233–239. doi:10.1080/k.1600-0412.1999.780311.x

Morgan, J. F., Lacey, J. H., & Sedgwick, P. M. (1999). Impact of pregnancy on bulimia nervosa. *British Journal of Psychiatry, 174*, 135. doi:10.1192/bjp.174.2.135

Morgan, J. F., Reid, F., & Lacey, J. H. (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *British Medical Journal, 319*, 1467-1468. doi:10.1136/bmj.319.7223.1467

Muazzam, A. B., & Khalid, R. (2011). Development and validation of Disordered Eating Behavior Scale: Identification, prevalence and difference with clinically diagnosed eating disorders. *Pakistan Journal of Psychological Research, 2*6, 127-148.

Muro-Sans P., Amador-Campos J. A., & Morgan J. F. (2008). The SCOFF-C: Psychometric properties of the Catalan version in a Spanish adolescent sample. *Journal of Psychosomatic Research, 64*, 81–86. doi:10.1016/j.jpsychores.2007.06.011

Murphy, M. K., Sanderson, C. F. B., Black, N. A., Askham, J., Lamping, D. L., & Marteau, T. (1998). Consensus development methods and their use in clinical guideline development. *Health Technology Assessment, 2,* 1–88.

Mustapic, J., Marcinko, D., & Vargek, P. (2015). Eating behaviours in adolescent girls: The role of body shame and body dissatisfaction. *Eating and Weight Disorders, 20,* 329-335. doi:10.1007/s40519-015-0183-2

National Eating Disorders Collaboration. (2017). *Disordered eating and dieting*. Retrieved from: http://www.nedc.com.au/disordered-eating

National Eating Disorders Collaboration. (2012). *Eating disorders in Australia*. Retrieved from: http://www.nedc.com.au/eating-disorders-in-australia

National Eating Disorders Collaboration. (2015). *Pregnancy and eating disorders: A professional's guide to assessment and referral.* Crows Nest, NSW: NEDC.

National Institute for Health and Clinical Excellence. (2017a). *Antenatal and postnatal mental health: The NICE guideline on clinical management and service guidance.* Manchester, London: NICE.

National Institute for Health and Clinical Excellence. (2004). *Eating disorders: Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. London: The British Psychological Society & Royal College of Psychiatrists.

National Institute for Health and Care Excellence. (2016). *Antenatal and postnatal mental health NICE unclassified quality standard (QS115).* Manchester, London: NICE.

National Institute for Health and Clinical Excellence. (2017b). *Eating disorders: Recognition and treatment.* Manchester, London: NICE.

Naylor, C., Das, P., Ross, S., Honeyman, M., Thompson, J., & Gilburt, H. (2016). *Bringing together physical and mental health: A new frontier for integrated care.* London: The King's Fund. Retrieved from: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Bringing-together-Kings-Fund-March-2016_1.pdf

Newton, M. S., & Chizawsky, L. L. (2006). Treating vulnerable populations: The case of eating disorders during pregnancy. *Journal of Psychosomatic Obstetrics & Gynecology, 27,* 5-7. doi:10.1080/01674820500451517

Noetel, M., Dawson, L., Hay, P., & Touyz, S. (2017). The assessment and treatment of unhealthy exercise in adolescents with anorexia nervosa: A Delphi study to synthesize clinical knowledge. *International Journal of Eating Disorders, 50,* 378-388. doi:10.1002/eat.22657

Norcross, J. C., Hedges, M., & Prochaska, J. O. (2002). The face of 2010: A Delphi poll on the future of psychotherapy. *Professional Psychology: Research and Practice, 33,* 316-322. doi:10.1037/0735-7028.33.3.316

Norcross, J. C., Koocher, G. P., & Garofalo, A. (2006). Discredited psychological treatments and tests: A Delphi poll. *Professional Psychology: Research and Practice, 37,* 515-522. doi:10.1037/0735-7028.37.5.515

Nunes, M., Pinheiro, A., Hoffmann, J., & Schmidt, M. (2012). Binge eating during pregnancy and birth outcomes: A cohort study in a disadvantages population in Brazil. *International Journal of Eating Disorders, 45,* 827-831. doi:10.1002/eat.22024

Nunes, M., Pinheiro, A., Hoffmann, J., & Schmidt, M. (2014). Eating disorders symptoms in pregnancy and postpartum: A prospective study in a disadvantaged population in Brazil. *International Journal of Eating Disorders, 47,* 426–430. doi:10.1002/eat.22236

Oates, M. (2015). *Perinatal mental health services. Recommendations for the provision of services for childbearing women*. London: Royal College of Psychiatrists. Retrieved from: www.rcpsych.ac.uk/usefulresources/publications/collegereports/cr/cr197.aspx

O'Hara, M. H., & Swain, A. M. (1996). Rates and risks of postpartum depression: A meta-analysis. *International Review of Psychiatry, 8,* 37-52. doi:10.3109/09540269609037816

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations, and applications. *Information and Management, 42,* 15-29. doi: 10.1016/j.im.2003.11.002

Olboni, C. M., & Alvarenga, M. S. (2015). Eating attitudes, attitudes related to weight gain, and body satisfaction of pregnant adolescents. *Revista Brasileira de Ginecologia e Obstetrícia, 37,* 585-592. doi:10.1590/S0100-720320150005481

Organisation for Economic Cooperation and Development. (2017). *OECD digital economy outlook 2017.* Retrieved from: http://www.oecd.org/internet/oecd-digital-economy-outlook-2017-9789264276284-en.htm

Orloff, N. C., & Hormes, J. M. (2014). Pickles and ice cream! Food cravings in pregnancy: Hypotheses, preliminary evidence, and directions for future research. *Frontiers in Psychology, 5,* 1-15. doi:10.3389/fpsycg.2014.01076

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies.* Princeton, NJ and Oxford: Princeton University Press.

Pannocchia L., Di Fiorino M., Giannini M., & Vanderlinden J. A. (2011). Psychometric exploration of an Italian translation of the SCOFF questionnaire. *European Eating Disorders Review, 19,* 371–373. doi:10.1002/erv.1105

Paré, E., Parry, S., McElrath, T. F., Pucci, D., Newton, A., & Lim, K. H., (2014). Clinical risk factors for preeclampsia in the 21st century. *Obstetrics and Gynecology, 124,* 763-770. doi:10.1097/AOG.0000000000000451

Park, R. J., Senior, R., & Stein, A. (2003). The offspring of mothers with eating disorders. *European Child and Adolescent Psychiatry, 12,* 110-119. doi:10.1007/s00787-003-1114-8

Parker, G., Fletcher, K., Barrett, M., Synnott, H., Breakspear, M., Hyett, M., & Hadzi-Pavlovic, D. (2008). Screening for bipolar disorder: The utility and comparative properties of the MSS and MDQ measures. *Journal of Affective Disorders, 109,* 83-89. doi:10.1016/j.jad.2007.11.003

Parker, S. C., Lyons, J., & Bonner, J. (2005). Eating disorders in graduate students: Exploring the SCOFF questionnaire as a simple screening tool. *Journal of American College Health, 54,* 103–107. doi:10.3200/JACH.54.2.103-107

Patel, P., Wheatcroft, R., Park, R. J., & Stein, A. (2002). The children of mothers with eating disorders. *Clinical Child and Family Psychology Review, 5,* 1-19. doi:10.1023/A:1014524207660

Patil, C. L. (2012). Appetite sensations in pregnancy among agropastoral women in rural Tanzania. *Ecology of Food and Nutrition, 51*, 431-443. doi:10.1080/03670244. 2012.696012

Paulson, J. F., & Bazemore, S. D. (2010) Prenatal and postpartum depression in fathers and its association with maternal depression: A meta-analysis *Journal of the American Medical Association, 303,* 1961–1969. doi:10.1001/jama.2010.605

Perry, L., Morgan, J., Reid, F., Brunton, J., O'Brien, A., Luck, A., & Lacey, A. (2002). Screening for symptoms of eating disorders: Reliability of the SCOFF screening tool with written compared to oral delivery. *International Journal of Eating Disorders, 32*, 466–472. doi:10.1002/eat.10093

Peterson, C. B., Crosby, R. D., Wonderlich, S. A., Joiner, T., Crow, S. J., Mitchell, J. E., Bardone-Cone, A. M., . . . & le Grange, D. (2007). Psychometric properties of the Eating Disorder Examination-Questionnaire: Factor structure and internal consistency. *International Journal of Eating Disorders, 40,* 386–389. doi:10.1002/eat.20373.

Pettersson, C. B., Zandian, M., & Clinton, D. (2016) Eating disorder symptoms pre- and postpartum. *Archives of Womens Mental Health, 19*, 675-680. doi:10/1007/s00737-016-0619-3

Piedmont, R. L. (2014). Inter-item correlations. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3303-3304). Dordrecht, Netherlands: Springer Netherlands.

Poobalan, A.S., Aucott, L.S., Ross, L., Smith, W.C., Helms, P.J., & Williams, J.H. (2007). Effects of treating postnatal depression on mother-infant interaction and child development: Systematic review. *British Journal of Psychiatry, 191,* 378-86. doi:10.1192/bjp.bp.106.032789

Pope, J., Skinner, J. D., & Carruth, B. R. (1992). Cravings and aversions of pregnant adolescents. *Journal of the American Dietetic Association, 92,* 1479-1482.

Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing, 41*, 376-382. doi:10.1046/j.1365-2648.2003.02537.x

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15. doi:10.1016/S0001-6918(99)00050-5

PricewaterhouseCoopers., & Centre of Perinatal Excellence. (2014). *Valuing perinatal mental health: A report of the latest cost estimates for Australian births in 2013*. Flemington, VIC: Centre of Perinatal Excellence.

Public Health England. (2013). *NHS population screening explained*. Retrieved from https://www.gov.uk/guidance/nhs-population-screening-explained

Quispel, C., Schneider, T., Hoogendijk, W., Bonsel, G. J., Lambregtse-van den Berg, M. P. (2015). Successful five-item triage for the broad spectrum of mental disorders in pregnancy: A validation study. *BMC Pregnancy and Childbirth, 15,* 51-62. doi:10.1186/s12884-015-0480-9

Reas, D. L., Grilo, C. M., & Masheb, R. M. (2006). Reliability of the Eating Disorder Examination-Questionnaire in patients with binge eating disorder. *Behaviour Research and Therapy, 44*, 43–51. doi:10.1016/j.brat.2005.01.004.

Reavley, N. J., Ross, A. M., Killackey, E., & Jorm, A. F. (2012). Development of guidelines to assist organisations to support employees returning to work after an episode of

anxiety, depression or a related disorder: A Delphi consensus study with Australian professionals and consumers. *BMC Psychiatry, 12*, 135. doi:10.1186/1471-244X-12-135

Reavley, N. J., Ross, A. M., Killackey, E., & Jorm, A. F. (2013). Development of guidelines for tertiary education institutions to assist them in supporting students with a mental illness: A Delphi consensus study with Australian professionals and consumers. *PeerJ, 1,* 43. doi:10.7717/peerj.43

Reay. R., Matthey, S., Ellwood, D., & Scott, M. (2011). Long-term outcomes of participants in a perinatal depression early detection program. *Journal of Affective Disorders, 129*, 94–103. doi:10.1016/j.jad.2010.07.035

Reid N. G. (1988). The Delphi technique: Its contribution to the evaluation of professional practice. In R. Ellis (Ed.), *Professional competence and quality assurance in the caring professions* (pp. 230-262). Kent, UK: Chapman and Hall.

Reid, A. J., Biringer, A., Carroll, J. D., Midmer, D., Wilson, L. M., Chalmers, B., & Stewart, D. E. (1998). Using the ALPHA form in practice to assess antenatal psychosocial health. Antenatal Psychosocial Health Assessment. *Canadian Medical Association Journal, 159,* 677–684.

Reilly, N., Harris, S., Loxton, D., Chojenta, C., Forder, P., Milgrom, J., & Austin, M-P. (2013a). Disparities in reported psychosocial assessment across public and private maternity settings: A national survey of women in Australia. *BMC Public Health, 13,* 632. doi:10.1186/1471-2458-13-632

Reilly, N., Harris, S., Loxton, D., Chojenta, C., Forder, P., Milgrom, J., & Austin, M-P. (2013b). Referral for management of emotional health issues during the perinatal period: Does mental health assessment make a difference? *Birth, 40,* 297-306. doi: doi:10.1111/birt.12067

Reilly, N., Harris, S., Loxton, D., Chojenta, C., Forder, P., Milgrom, J., & Austin, M-P. (2014). The impact of routine assessment of past or current mental health on help-seeking in the perinatal period. *Women and Birth, 27,* 20-27. doi:10.1016/j.wombi.2014.07.003

Reneman, M. F., Dijkstra, A., Geertzen, J. H., & Dijkstra, P. U. (2010). Psychometric properties of Chronic Pain Acceptance Questionnaires: A systematic review. *European Journal of Pain, 14,* 457-465. doi:10.1016/j.ejpain.2009.08.003

Rø, Ø, Reas, D. L., & Stedal, K. (2015). Eating Disorder Examination Questionnaire (EDE-Q) in Norwegian adults: Discrimination between female controls and eating disorder patients. *European Eating Disorders Review, 23,* 308-412. doi:10.1002/erv.2372

Rocco, P. L., Orbitello, B., Perini, L., Pera, V., Ciano, R. P., & Balestrieri, M. (2005). Effects of pregnancy on eating attitudes and disorders: A prospective study. *Journal of Psychosomatic Research, 59,* 175-179. doi:10.1016/j.jpsychores.2005.03.002

Rollans, M., Schmied, V., Kemp, L., & Meade, T. (2013). Digging over that old ground: An Australian perspective of women's experience of psychosocial assessment and depression screening in pregnancy and following birth. *BMC Women's Health, 13*, 18. doi:10.1186/1472-6874-13-18

Roseboom, T. J., van der Meulen, J. H., Ravelli, A. C., Osmond, C., Barker, D. J., & Bleker, O. P. (2001). Effects of prenatal exposure to the Dutch famine on adult disease in later life: An overview. *Molecular and Cellular Endocrinology, 185*, 93–98. doi: 10.1016/S0303-7207(01)00721-3

Ross, A. M., Kelly, C. M., & Jorm, A. F. (2014a). Re-development of mental health first aid guidelines for non-suicidal self-injury: A Delphi study. *BMC Psychiatry, 14,* 236.

Ross, A. M., Kelly, C. M., & Jorm, A. F. (2014b). Re-development of mental health first aid guidelines for suicidal ideation and behaviour: A Delphi study. *BMC Psychiatry, 14*, 241. doi:10.1186/s12888-014-0236-5

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15*, 353-355. doi:10.1016/S0169-2070(99)00018-7

Rueda G. E., Díaz L. A., Campo A., Barros J. A., Avila G. C., Oróstegui L. T., . . . & Cadena L. P. (2005). Validation of the SCOFF questionnaire for screening of eating disorders in university women. *Biomédica, 25*, 196–202.

Rueda G. E., Díaz L. A., Ortiz D. P., Pinzón C., & Rodrigues J. (2005). Validation of the SCOFF questionnaire for screening the eating behavior disorders of adolescents in school. *Atención Primaria, 35*, 89–94.

Rusner, M., Berg, M., & Begley, C. (2016). Bipolar disorder in pregnancy and childbirth: A systematic review of outcomes. *BMC Pregnancy and Childbirth, 16*, 331. doi:10.1186/s12884-016-1127-1

Sackett, D. L., & Rosenberg, W. M. (1995). The need for evidence-based medicine. *Journal of the Royal Society of Medicine, 88*, 620–624. doi:10.1002/hec.4730040401

Santos, A. M., Benute, G. R. G., Nomura, R. M. Y., Santos, N. O., De Lucia, M. C. S., & Francisco, R. P. V. (2016). Pica and eating attitudes: A study of high-risk pregnancies. *Maternal and Child Health Journal, 20*, 577-582. doi:10.1007/s10995-015-1856-1

Santos, A. M., Benute, G. R. G., Santos, N. O., Nomura, R. M. Y., De Lucia, M. C. S., & Francisco, R. P. V. (2017). Presence of eating disorders and its relationship to anxiety and depression in pregnant women. *Midwifery, 51*, 12-15. doi: 10.1016/j.midw.2017.05.005

Sawyer, S. M., Tully, M. M., & Colin, A. A. (2001). Reproductive and sexual health in males with cystic fibrosis: A case for health professional education and training. *Journal of Adolescent Health, 28*, 36-40. doi:10.1016/S1054-139X(00)00172-5

Schneider, Z. (2002). An Australian study of women's experiences of their first pregnancy. *Midwifery, 18*, 238-249. doi:10.1054/midw.2002.0309

Sedgwick, P. (2011). Screening tests: Likelihood ratios. *British Medical Journal, 342*, 398. doi:10.1136/bmj.d3986

Senior, R., Barnes, J., & Emberson, J. R. (2005). Early experiences and their relationship to maternal eating disorder symptoms, both lifetime and during pregnancy. *British Journal of Psychiatry, 187,* 268-273. doi:10.1192/bjp.187.3.268

Shakespeare, J., Blake, F., & Garcia, J. A. (2003). Qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale. *British Journal of General Practice, 53*, 614–619.

Sharp, C. W., & Freeman, C. P. (1993). The medical complications of anorexia nervosa. *British Journal of Psychiatry, 162*, 452-462. doi:10.1192/bjp.162.4.452

Shisslak, C., Crago, M., & Estes, L. (1995). The spectrum of eating disturbances. *International Journal of Eating Disorders, 18*, 209-219. doi:10.1002/1098-108x(199511)18:3<209::aid-eat2260180303>3.0.co;2-e

Siega-Riz, A. M., Evenson, K. R., & Dole, N. (2004). Pregnancy-related weight gain: A link to obesity? *Nutrition Reviews, 62*, 105–111. doi:10.1301/nr.2004.jul.S105-S111

Siega-Riz, A. M., Haugen, M., Meltzer, H. M., Von Holle, A., Hamer, R., Torgersen L., . . . & Bulik, C. M. (2008). Nutrient and food group intakes of women with and without bulimia nervosa and binge eating disorder during pregnancy. *American Journal of Clinical Nutrition, 87*, 1346-1355.

Siega-Riz, A. M., Von Holle, A., Haugen, M., Meltzer, H. M., Hamer, R., Torgersen, L., . . . & Bulik, C. M. (2011). Gestational weight gain of women with eating disorders in the Norwegian pregnancy cohort. *International Journal of Eating Disorders, 44,* 428-434. doi:10.1002/eat.20835

Sinclair, D., & Murray, L. (1998). Effects of postnatal depression on children's adjustment to school. Teacher's reports. *British Journal of Psychiatry, 172*, 58-63. doi:10.1192/bjp.172.1.58

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Medicine, 8*, e1000393. doi:10.1371/journal.pmed.1000393

Skouteris, H., Carr, R., Wertheim, E.H., Paxton, S.J., & Duncombe, D. (2005). A prospective study of factors that lead to body image dissatisfaction during pregnancy. *Body Image, 2,* 347–361. doi:10.1016/j.bodyim.2005.09.002

Slane, J. D., & Levine, M. D. (2015). Association of restraint and disinhibition to gestational weight gain among pregnant former smokers. *Women's Health Issues, 25,* 300-305. doi:10.1016/j.whi.2015.03.005

Smink, F. R., van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: Incidence, prevalence, and mortality rates. *Current Psychiatry Reports, 14,* 406-414. doi:10.1007/s11920-012-0282-y

Smink, F. R., van Hoeken, D., Oldehinkel, A. J., & Hoek, H. W. (2014). Prevalence and severity of DSM-5 eating disorders in a community cohort of adolescents. *International Journal of Eating Disorders, 47,* 610-619. doi:10.1002/eat.22316

Smith, C. A., Grant, S., Lyttleton, J. & Cochrane, S. (2012). Using a Delphi consensus process to develop an acupuncture treatment protocol by consensus for women

undergoing Assisted Reproductive Technology (ART). *BMC Complementary and Alternative Medicine, 12,* 88. doi:10.1186/1472-6882-12-88

Soares, R. M., Nunes, M. A., Schmidt, M. A., Giacomelle, A., Manzolli, P., Camey, S., Buzz, C., . . . & Duncan, B. B. (2009). Inappropriate eating behaviours during pregnancy: Prevalence and associated factors among pregnant women attending primary care in southern Brazil. *International Journal of Eating Disorders, 42,* 387-393. doi:10.1002/eat.20643

Sohail, R., & Muazzam, A. (2012). Correlates of disordered eating behaviour among pregnant women. *Pakistan Journal of Psychological Research, 27,* 153-172.

Solmi, F., Hatch, S. Hotopf, M., Treasure, J., & Micali, N. (2014). Prevalence and correlates of disordered eating in a general population sample: The South East London Community Health (SELCoH) study. *Social Psychiatry and Psychiatric Epidemiology, 49,* 1335-1346. doi:10.1007/s00127-014-0822-3

Solmi, F., Hatch, S. Hotopf, M., Treasure, J., & Micali, N. (2015). Validation of the SCOFF questionnaire for eating disorders in a multiethnic general population. *International Journal of Eating Disorders, 48,* 312-316. doi:10.1002/eat.22373

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15,* 201-293. doi:10.2307/1412107

Spinelli, E. (1983). *The interpreted world: An introduction to phenomenological psychology.* Thousand Oaks, CA: Sage Publications.

Spitzer, R. L., Williams, J. B., Kroenke, K., Homyak, R., & McMurray, J. (2000). Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American Journal of Obstetrics and Gynecology, 183,* 759–769. doi:10.1067/mob.2000.106580

Squires, C., Lalanne, C., Murday, N., Simoglou, V., & Vaivre-Douret, L. (2015). The influence of eating disorders on mothers' sensitivity and adaptation during feeding: A longitudinal observational study. *BMC Pregnancy and Childbirth, 14,* 274-283. doi:10.1186/1471-2393-14-274

Stein, A., & Fairburn, C. G. (1996). Eating habits and attitudes in the postpartum period. *Psychosomatic Medicine, 4,* 321-325. doi:10.1097/00006842-199607000-00004

Stewart, D. E., Robinson, G. E., Goldbloom, D. S., & Wright, C. (1990). Infertility and eating disorders. *American Journal of Obstetrics and Gynecology, 163,* 1196-1199. doi:10.1016/0002-9378(90)90688-4

Stice, E., Fisher, M., & Martinez, E. (2004). Eating Disorder Diagnostic Scale: Additional evidence of reliability and validity. *Psychological Assessment, 16,* 60. doi:10.1037/1040-3590.16.1.60

Stice, E., Marti, C. N., & Rohde, P. (2013). Prevalence, incidence, impairment, and course of the proposed dsm-5 eating disorder diagnoses in an 8-year prospective community study of young women. *Journal of Abnormal Psychology, 122,* 445-457. doi:10.1037/a0030679

Stice, E., Shaw, H.E., & Marti, C. (2007). A meta-analytic review of eating disorder prevention programs: Encouraging findings. *Annual Review of Clinical Psychology, 3,* 207-231. doi:10.1146/annurev.clinpsy.3.022806.091447

Smolak, L., & Striegel-Moore, R. H. (2001). Challenging the myth of the golden girl. In R. H. Striegel-Moore & L. Smolak (Eds.), *Eating disorders: Innovative directions in research and practice* (pp. 111-132). Washington, D. C.: American Psychological Association.

Stein, A., Pearson, R. M., Goodman, S. H., Rapa, E., Rahman, A., McCallum, M., Howard, L. M., & Pariante, C. M. (2014). Effects of perinatal mental disorders on the fetus and child. *The Lancet, 384,* 1800–1819. doi:10.1016/S0140-6736(14)61277-0

Strimling, B. (1984). Infant of a pregnancy complicated by anorexia nervosa. *American Journal of Diseases of Children, 138,* 68–69. doi:10.1001/archpedi.1984.02140390056017

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research, 29*, 71–83. doi:10.1016/0022-3999(85)90010-8

Sullivan, P. F., Bulik, C. M., Fear, J. L., & Pickering, A. (1998). Outcome of anorexia nervosa: A case-control study. *American Journal of Psychiatry, 155*, 939-946. doi:10.1176/ajp.155.7.939

Sullivan, E. L., Grayson, B., Takahashi, D., Robertson, N., Maier, A., Bethea, C. L., . . . & Grove, K. L. (2010). Chronic consumption of a high-fat diet during pregnancy causes perturbations in the serotonergic system and increased anxiety-like behavior in nonhuman primate offspring. *Journal of Neuroscience, 30,* 3826-3830. doi:10.1523/JNEUROSCI.5560-09.2010

Sumison, T. (1998). The Delphi technique: An adaptive research tool. *British Journal of Occupational Therapy, 61,* 153-156. doi:10.1177/030802269806100403

Sumner, A., Waller, G., Killick, S., & Elstein, M. (1993). Body image distortion in pregnancy: A pilot study of the effects of media images. *Journal of Reproductive and Infant Psychology, 11,* 203-208. doi:10.1080/02646839308403219

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few*. London: Abacus.

Svensson, E. D. (2014). Validity of scales. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1637-1639). Dordrecht, Netherlands: Springer Netherlands.

Swanson, S. A., Crow, S. J., Le Grange, D., Swendsen, J., & Merikangas, K. R. (2011). Prevalence and correlates of eating disorders in adolescents. Results from the national comorbidity survey replication adolescent supplement. *Archives of General Psychiatry, 68,* 714-723. doi:10.1001/archgenpsychiatry.2011.22

Sword ,W., Busse, D., Ganann, R., McMillan, T., & Swinton, M. (2008). Women's care-seeking experiences after referral for postpartum depression. *Qualitative Health Research, 18,* 1161–1173. doi:10.1177/1049732308321736

Sysko, R., Walsh, B. T., & Fairburn, C. G. (2005). Eating Disorder Examination-Questionnaire as a measure of change in patients with bulimia nervosa. *International Journal of Eating Disorders, 37,* 100–106. doi:10.1002/eat.20078

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60,* 34-42. doi:10.1016/j.jclinepi.2006.03.012

Thelen, M.H., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the Bulimia Test: The BULIT-R. *Psychological Assessment, 3,* 119-124. doi:10.1037//1040-3590.3.1.119

Thielsen, V. A., & Leahy, M. J. (2001). The Delphi method in rehabilitation counseling. *Research Rehabilitation Counseling Bulletin, 44,* 196-208. doi:10.1177/003435520104400402

Thomas, N., Grunnet, L. G., Poulsen, P., Christopher, S., Spurgeon, R., Inbakumari, M., . . . & Bygbjerg, I. C. (2012). Born with low birth weight in rural Southern India: What

are the metabolic consequences 20 years later? *European Journal of Endocrinology, 166,* 647–655. doi:10.1530/EJE-11-0870

Thomas, J. J., Vartanian, L. R., & Brownell, K. D. (2009). The relationship between eating disorder not otherwise specified (EDNOS) and officially recognized eating disorders: meta-analysis and implications for DSM. *Psychological Bulletin, 135,* 407–433. doi:10.1037/a0015326

Thompson, B. (2003). *Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage Publications.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of the mind*. Chicago: Chicago University Press.

Tierney, S., Fox, J. R. E., Butterfield, C., Stringer, E. & Furber, C. (2011). Treading the tight-rope between motherhood and an eating disorder: A qualitative study. *International Journal of Nursing Studies, 48,* 1223–1233. doi:10.1016/j.ijnurstu.2010.11.007

Tierney, S., McGlone, C., & Furber, C. (2013). What can qualitative studies tell us about the experiences of women who are pregnant that have an eating disorder? *Midwifery, 29,* 542-549. doi:10.1016/j.midw.2012.04.013

Tierson, F. D., Olsen, C. L., & Hook, E. B. (1985). Influence of cravings and aversions on diet in pregnancy. *Ecology of Food and Nutrition, 17,* 117-129. doi:10.1080/0367 0244.1985.9990886

Tiller, J., & Treasure, J. (1998). Eating disorders precipitated by pregnancy. *European Eating Disorders Review, 6,* 178-187. doi:10.1002/(SICI)1099-0968(199809).x

Torgersen, L., Von Holle, A., Reichborn-Kjennerud, T., Knoph Berg, C., Hamer, R., Sullivan, P., & Bulik, C. M. (2008). Nausea and vomiting of pregnancy in women with bulimia nervosa and eating disorders not otherwise specified. *International Journal of Eating Disorders, 41,* 722-727. doi:10.1002/eat.20564

Torgersen, L., Ystrom, E., Haugen, M., Meltzer, H. M., Von Holle, A., Knoph Berg, C., Reichborn-Kjennerud, T., & Bulik, C. M. (2010). Breastfeeding practice in mothers with eating disorders. *Maternal and Child Nutrition, 6,* 243-252. doi:10.1111/j.1740-8709.2009.00208.x

Treasure, J., & Russell, G. (2011). The case for early intervention in anorexia nervosa: Theoretical exploration of maintaining factors. *British Journal of Psychiatry, 199,* 5-7. doi:10.1192/bjp.bp.110.087585

Tremblay, K. A. (2015). *Eating and psychological distress during pregnancy: The use of ecological momentary assessment* (doctoral dissertation). Retrieved from: ProQuest Dissertations and Theses database (UMI No. 3732671).

Trevelyan, E. G., & Robinson, N. (2015). Delphi methodology in health research: How to do it? *European Journal of Integrative Medicine, 7,* 423-428. doi:10.1016/j.eujim.2015.07.002

Turner, H., & Bryant-Waugh, R. (2004). Eating disorder not otherwise specified (EDNOS): Profiles of clients presenting at a community eating disorder service. *European Eating Disorders Review, 12,* 18-26. doi:10.1002/erv.552

Turton, P., Hughes, P., Bolton, H., & Sedgwick, P. (1999). Incidence and demographic correlates of eating disorder symptoms in a pregnant population. *International Journal of Eating Disorders, 26,* 448-452. doi:10.1002/(SICI)1098-108X(199912)26:4%3C448::AID-EAT10%3E3.0.CO;2-3

Ulfvebrand, S., Birgegård, A., Norring, C., Högdahl, L., & von Hausswolff-Juhlin, Y. (2015). Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. *Psychiatry Research, 230,* 294-299. doi:10.1016/j.psychres.2015.09.008

United Nations. (2014). *The millennium development goals report 2014*. New York: United Nations.

van Strien, T., Frijters, J. E., Bergers, G. P., & Defares, P. B. (1986). The Dutch Eating Behaviour Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behaviour. *International Journal of Eating Disorders, 5*, 295-315. doi:10.1002/1098-108X(198602)5:2<295::AID-EAT2260050209>3.0.CO;2-T

Verbeek, T., Bockting, C. L., van Pampus, M. G., Ormel, J., Meijer, J. L., Hartmen, C. A., & Burger, H. (2012). Postpartum depression predicts offspring mental health problems in adolescence independently of parental lifetime psychopathology. *Journal of Affective Disorders, 136*, 948-954. doi:10.1016/j.jad.2011.08.035

von der Gracht, H. A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change, 79*, 1525–1536. doi:10.1016/j.techfore.2012.04.013

Wade, T. D., Bergin, J. L., Tiggemann, M., Bulik, C. M., & Fairburn, C. G. (2006). Prevalence and long-term course of lifetime eating disorders in an adult twin cohort. *Australian and New Zealand Journal of Psychiatry, 40*, 121-128. doi:10.1111/j.1440-1614.2006.01758.x

Wade, T. D., Byrne, S., & Bryant-Waugh, R. (2008). The Eating Disorder Examination: Norms and construct validity with young and middle adolescent girls. *International Journal of Eating Disorders, 41*, 551–558. doi:10.1002/eat.20526

Wade, T. D., Crosby, R. D., & Martin, N. G. (2006). Use of latent profile analysis to identify eating disorder phenotypes in an adult Australian twin cohort. *Archives of General Psychiatry, 63*, 1377-1384. doi:10.1001/archpsyc.63.12.1377

Walker A.M. & Selfe J. (1996) The Delphi method: A useful tool for the allied health researcher. *British Journal of Therapy and Rehabilitation 3,* 677-681. doi:10.12968/bjtr.1996.3.12.14731

Wallace, K. (2013). *'Pregorexia': Extreme dieting while pregnant*. Retrieved from: http://edition.cnn.com/2013/11/20/living/pregnant-dieting-pregorexia-moms/

Walsh, T. B., & Sysko, R. (2009). Broad categories for the diagnosis of eating disorders (BCD-ED): An alternative system for classification. *International Journal of Eating Disorders, 42,* 754-764. doi:10/1002/eat.20722

Ward, V. B. (2008). Eating disorders in pregnancy. *British Medical Journal, 336,* 93-96. doi:10.1136/bmj.39393.689595.BE

Ward, L., Stebbings, S., Sherman K. J., Cherkin, D., & Baxter, G. D. (2014). Establishing key components of yoga interventions for musculoskeletal conditions: A Delphi survey. *BMC Complementary and Alternative Medicine, 14,* 196. doi:10.1186/1472-6882-14-196.

Watkins, B., Cooper, P., & Lask, B. (2011). History of eating disorder in mothers of children with early onset eating disorder or disturbance. *European Eating Disorders Review, 20,* 121-125. doi:10.1002/erv.1125

Watkins, B., Cooper, P., & Lask, B. (2011). History of eating disorder in mothers of children with early onset eating disorder or disturbance. *European Eating Disorders Review, 20,* 121-125. doi:10.1002/erv.1125

Watson, H. J., Torgersen, L., Zerwas, S., Reichborn-Kjennerud, T., Knoph, C., Stoltenberg, C., Siega-Riz, A. M., . . . & Bulik, C. M. (2014). Eating disorders, pregnancy, and the postpartum period: Findings from the Norwegian Mother and Child Cohort Study (MoBa). *Norsk Epidemiologi, 24,* 51-62. doi:10.5324/nje.v24i1-2.1758

Watson, H. J., Von Holle, A., Hamer, R. M., Knoph Berg, C., Torgersen, L., Magnus, P., Stoltenberg, C., . . . & Bulik, C. M. (2012). Remission, continuation and incidence of eating disorders during pregnancy: A validation study in a population-based pregnancy cohort. *Psychological Medicine, 43,* 1723-1734. doi:10.1017/S0033291712002516

Waugh, E., & Bulik, C. M. (1999). Offspring of women with eating disorders. *International Journal of Eating Disorders, 25,* 123-133. doi:10.1002/(SICI)1098-108X(199903)25:2%3C123::AID-EAT1%3E3.0.CO;2-B

Weinfeld, R., Dubay, M., Burchell, R., Mellerick, J. & Kennedy, A. (1977). Pregnancy associated with anorexia and starvation. *American Journal of Obstetrics and Gynecology, 15,* 698–699.

Wiles, R. (1994). I'm not fat, I'm pregnant. In S. Wilkinson & C. Kitzinger (Eds.),*Women and health: Feminist perspectives* (pp. 33-48). London: Taylor Francis.

Wiles, R. (1998). The views of women of above average weight about appropriate weight gain in pregnancy. *Midwifery, 14*, 254-260. doi:10.1016/S0266-6138(98)90098-5

Wilfley, D. E., Schwartz, M. B., Spurrell, E. B., & Fairburn, C. G. (2000). Using the Eating Disorder Examination to identify the specific psychopathology of binge eating disorder. *International Journal of Eating Disorders, 27*, 259–269. doi:10.1002/(SICI)1098-108X(200004)27:3<259::AID-EAT2>3.0.CO;2-G

Williams, M., & Haverkamp, B. E. (2010). Identifying critical competencies for psychotherapeutic practice with eating disordered clients: A Delphi study. *Eating Disorders, 18,* 91–109. doi:10.1080/10640260903585524

Wilson, J. M. G., & Jungner., G. (1968). *Principles and practices of screening for disease.* Geneva, Switzerland: World Health Organisation.

Wisner, K. L., Sit, D. K., McShea, M. C., Rizzo, D. M., Zoretich, R. A., Hughes, C. L., Eng, H. F., . . . & Hanusa, B. H. (2013). Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen positive depression findings. *JAMA Psychiatry, 70,* 490–498. doi:10.1001/jamapsychiatry.2013.87

Wolfe, B. E. (2005). Reproductive health in women with eating disorders. *JOGNN, 34,* 255-263. doi:10.1177/0884217505274595

World Health Organisation. (1992). *ICD-10 classifications of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: WHO.

World Health Organisation. (2009). *Maternal mental health and child health and development in resource-constrained settings*. Geneva, Switzerland: WHO.

World Health Organisation. (2016). *WHO recommendations on antenatal care for a positive pregnancy experience.* Geneva, Switzerland: WHO.

Wren, C. (2011). Prematurity, low birth weight, and cardiovascular malformations. *Pediatrics, 127,* 385–386. doi:10.1542/peds.2010-3345

Yap, M. B., Pilkington, P. D., Ryan, S. M., Kelly, C. M., & Jorm, A. F. (2014). Parenting strategies for reducing the risk of adolescent depression and anxiety disorders: A Delphi consensus study. *Journal of Affective Disorders, 156,* 67–75. doi:10.1016/ j.jad.2013.11.017

Yawn, B. P., Dietrich, A. J., Wollan, P., Bertram, S., Graham, D., Huff, J., Kurland, M., . . . & TRIPPD Practices. (2012). TRIPPD: A practice-based network effectiveness study of postpartum depression screening and management. *Annals of Family Medicine, 10,* 320–329. doi:10.1370/afm.1418

Verbeek, T., Bockting, C. L., van Pampus, M., Ormel, J., Meijer, J. L., Hartman, C. A., & Burger, H. (2012). Postpartum depression predicts offspring mental health problems

in adolescence independently of parental lifetime psychopathology. *Journal of Affective Disorders, 136*, 948–954. doi:10.1016/j.jad.2011.08.035

Veritas Health Innovation. (2018). *Covidence systematic review software*. Melbourne, Australia. Available at: www.covidence.org

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*, 561-577.

Appendix A

Ethics Approval for Studies 1 and 2

BOND UNIVERSITY
BRINGING AMBITION TO LIFE

HUMAN RESEARCH
ETHICS COMMITTEE
Bond University
Gold Coast, Queensland 4229
Australia
Ph: +61 7 5595 4194
Fax: +61 7 5595 1528
(from overseas)
Email: ethics@bond.edu.au
ABN 88 010 694 121
CRICOS CODE 00017B

11 August 2015

Peta Stapleton, Amy Bannatyne, Roger Hughes and Kristen MaKenzie-Shalders
Faculty of Society and Design
Bond University

Dear Peta

Application ID: **15278**
Project Title: **How do we define and screen for disordered eating and eating disorders in pregnancy? A delphi study to understand expert perspectives**

I am pleased to confirm that your project was reviewed by Bond University Human Research Ethics Committee and you have been granted approval to proceed.

The Committee requires, as a condition of approval, that all investigations be carried out in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007). Approval is subject to conduct of the research in accordance with the requirements set out in the National Statement.

Approval is given subject to the protocol of the study being undertaken as described in your application, and approved amendments. As you may be aware the Ethics Committee is required to annually report on the progress of research it has approved. We would greatly appreciate if you could respond promptly and fully to the request for information on this project which will be distributed in March/April each year.

Under the terms of the National statement BUHREC has a role to monitor approved research projects and if necessary may withdraw approval. Conduct of unapproved research or deviation from the approved protocol may constitute academic misconduct and will be investigated in accordance with Section B of the *Australian Code for the Responsible Conduct of Research* (2007). Please refer to the Research Ethics website for more detail on Research Integrity and Bond University processes for dealing with instances of research misconduct.

You are reminded that the Principal Investigator must immediately report anything that might warrant review of ethical approval of the project. Should you have any queries or experience any problems, please contact us promptly.

We wish you well with your research project.

Yours sincerely

Dr Mark Bahr
Chair Bond University Human Research Ethics Committee

Appendix B

Round I Delphi Questionnaire – Professional Panel

PARTICIPANT INFORMATION SHEET

How do we define and screen for disordered eating in pregnancy? A Delphi study to understand expert perspectives.

Ethics Protocol Number: 0000015278

Principal Investigators: Dr Peta Stapleton[1], Prof Roger Hughes[2,3], Dr Bruce Watt[1], and Dr Kristen MacKenzie-Shalders[2]
HDR Student Investigator: Ms Amy Bannatyne[1]

[1] School of Psychology, Bond University, Robina, Queensland
[2] Faculty of Health Sciences & Medicine, Bond University, Robina, Queensland
[3] Office of Research Services, Bond University, Robina, Queensland

You are invited to participate in a modified Delphi study exploring expert practitioner views on disordered eating in the antenatal period. Specifically, the expression, distinction, and assessment of disordered eating in pregnancy. This research is being conducted by Ms Amy Bannatyne, a PhD candidate in the School of Psychology at Bond University, under the supervision of Dr Peta Stapleton, Professor Roger Hughes, Dr Bruce Watt and Dr Kristen MacKenzie-Shalders. It is hoped this study will clarify the signs and symptoms of disordered eating in pregnancy, in addition to providing reliable and professional guidance to assist in the development a pregnancy-specific screening tool for subclinical eating disorder symptomatology (disordered eating) in antenatal care.

Your participation in this expert panel involves responding to a series of structured questionnaires, administered over several stages or 'rounds'. During these rounds you will be asked to respond to qualitative and quantitative items, enabling you to express and explain your views and perspectives based on your experiences and research. It is anticipated each questionnaire will take approximately 30 minutes of your time. Information from each round will be used to progressively refine or develop new questions for subsequent rounds. At the commencement of each new round, you will be provided with a summary of results/responses from the previous round. This iterative process will continue until consensus among the group is reached, which is typically achieved within 3 rounds. We anticipate this will take about 6 months.

It is important to note your participation in this expert panel will be completely confidential. While your identity will be known to the panel moderator (Ms Bannatyne), other panellists will not have knowledge of your identity. Anonymity is vital for true consensus to be reached, free from bias and peer influence. Your responses will also remain completely confidential following the conclusion of

the study. Data will be stored in a secure location at Bond University for a period of five years in accordance with the guidelines outlined by the Bond University Human Research Ethics Committee. Your participation is voluntary; therefore, you can withdraw from the study at any time.

Should you have any complaints concerning the manner in which this research project is conducted, please do not hesitate to contact the Bond University Research Ethics Committee:

Bond University Human Research Ethics Committee
Office of Research Services
Bond University, Gold Coast, 4229 Australia
Tel: +61 7 5595 4194 Fax: +61 7 5595 1120
Email: ethics@bond.edu.au

We thank you for taking time to assist us with this research. If you would like further information about the project, please contact Ms Amy Bannatyne at abannaty@bond.edu.au.

International Consensus on the Expression, Distinction, and Assessment of Disordered Eating in Pregnancy

ROUND I QUESTIONNAIRE

Round I Delphi Procedure

- The current study has employed a modified Delphi approach, whereby Round I (this questionnaire) has been pre-populated with items following a systematic literature search. We do, however, encourage you to suggest additional items for consideration in Round II using the open-ended text boxes throughout the questionnaire.

- This questionnaire is comprised of three sections:
    1. Possible signs and symptoms of disordered eating in pregnancy
        - In this section, you will be asked to indicate the extent to which you agreed that an item reflects a sign or symptom of disordered eating in pregnancy on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).
    2. Possible factors (foci) that may help distinguish disordered eating symptomatology from pregnancy-appropriate symptomatology/changes
        - In this section, you will be asked to indicate how important certain factors are in distinguishing disordered eating symptomatology from pregnancy-appropriate symptomatology (foci items) on a 5-point Likert scale (1 = *not important* to 5 = *very important*).
    3. Perceptions of assessment
        - In this section, you will be asked to indicate whether screening for disordered eating should be a routine component of antenatal care (i.e., occur for every woman), only occur when indicated by presenting signs/symptoms and/or historical factors, or not occur at all. You will be asked to rate each option on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). You will also be asked to review and rate the suitability of 12 potential assessment methods for identifying disordered eating in antenatal care. These methods were again rated on a 5-point Likert scale (1 = *not suitable at all* to 5 = *very suitable*).

- Please do not hesitate to contact me via email (abannaty@bond.edu.au) if you have any queries.

DEMOGRAPHIC QUESTIONS

Please answer the following questions so we can learn a little more about you and your expertise. Your answers will not be shared with the group in a way that would reveal your identity.

Please indicate your sex:

☐ Male
☐ Female

Please indicate your age in years: _____

Please indicate which country you currently reside/practice: _____

Please indicate your highest level of education:

☐ Bachelor degree
☐ Masters degree
☐ Doctorate or PhD
☐ Other (please specify) _____

Please indicate your professional area:

☐ Psychology
☐ General medicine
☐ Psychiatry
☐ Obstetrics
☐ Nursing or midwifery
☐ Nutrition and dietetics
☐ Social work
☐ Other (please specify) _____

Please indicate how many years you have worked in your profession: _____

Please indicate how many years you have been involved in the field of disordered eating (clinical practice and/or research): _____

Please select the best description of your current professional activities:

☐ Clinical practice (no research activities)
☐ Researcher (no clinical practice)
☐ Researcher who is also involved in clinical practice
☐ Clinician who is also involved in some research
☐ Retired
☐ Other (please specify) _____

Are you a Fellow of the Academy of Eating Disorders? _____

SECTION 1 – POSSIBLE SIGNS & SYMPTOMS OF DISORDERED EATING IN PREGNANCY

Below is a list of possible signs and symptoms of disordered eating in pregnancy populated from existing literature. Please rate the extent to which you agreed that an item reflects a sign or symptom of disordered eating in pregnancy on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). Currently, the list is quite lengthy; however, it is important we consider all possible attributes to increase the likelihood of a clear definition. We anticipate the list will reduce in subsequent rounds. Please note, if you have any comments regarding the attributes or if we have missed any you believe are important and need to be included for consideration in Round II, please leave your feedback in the comment box located below the list.

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
The woman's dietary consumption does not appear sufficient to support a healthy pregnancy	1	2	3	4	5
The woman reports dieting behaviours (e.g., calorie counting)	1	2	3	4	5
The woman reports (or appears to have) inflexibility and rigidity with diet (i.e., strict consumption of diet foods only)	1	2	3	4	5
The woman reports fasting and/or skipping meals	1	2	3	4	5
The woman reports using meal replacements (when not advised by a health professional) for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports she is repeatedly weighing herself (e.g., every day or several times a day) to monitor gestational weight gain	1	2	3	4	5
The woman reports she is unable to eat outside of her home for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports eating in secret because she is ashamed of the amount of food she has consumed and/or her pregnancy weight/shape/size	1	2	3	4	5
The woman reports eating an objectively large amount of food and feeling distressed/disgusted by this	1	2	3	4	5
The woman reports she has been eating for "two"	1	2	3	4	5
The woman reports she is eating when not physically hungry	1	2	3	4	5
The woman reports using food to cope with/soothe strong emotions, or to reward herself	1	2	3	4	5
The woman reports eating rapidly and until uncomfortably full	1	2	3	4	5
The woman reports she has been engaging in self-induced vomiting to control her pregnancy weight/shape/size	1	2	3	4	5
The woman reports she has been obsessively exercising for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports she has been exercising against medical recommendations for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The woman reports she has been exercising in secret for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports she refuses to purchase maternity clothing	1	2	3	4	5
The woman reports wearing specific clothing to conceal pregnancy	1	2	3	4	5
The appears to be (or reports) misusing gestational diabetes medication for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports using laxatives or enemas to reduce gestational weight gain/induce weight loss	1	2	3	4	5
The woman reports using appetite suppressants or “diet pills” for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports using natural supplements (e.g., tea detoxes) for the purpose of controlling pregnancy weight/shape/size	1	2	3	4	5
The woman reports (or appears to be) engaging in body checking behaviours to monitor pregnancy related changes in weight/shape/size	1	2	3	4	5
The woman reports incidences of self-harm due to distress about changes in weight/shape/size	1	2	3	4	5
The woman has a low body weight during pregnancy	1	2	3	4	5
The woman is losing weight while pregnant	1	2	3	4	5
The woman has not gained an adequate amount of weight during pregnancy based on her pre-pregnancy BMI	1	2	3	4	5
The woman has gained an excess amount of weight during pregnancy based on her pre-pregnancy BMI	1	2	3	4	5
The woman is experiencing rapid gestational weight gain	1	2	3	4	5
The woman reports dizziness and/or fatigue during pregnancy	1	2	3	4	5
The woman reports feeling nauseated most of the time during pregnancy	1	2	3	4	5
The woman is suffering from severe morning sickness that does not stop after the first trimester (hyperemesis gravidarum)	1	2	3	4	5
The woman reports or presents with dehydration	1	2	3	4	5
The woman reports or presents with abdominal bloating	1	2	3	4	5
The woman reports or presents with gastrointestinal discomfort	1	2	3	4	5
The woman reports her pregnancy body shape, size, or weight to be extremely important in determining her self-worth and importance (i.e., overvaluation of body shape and weight)	1	2	3	4	5
The woman appears to perceive herself to be overweight for pregnancy stage, when objectively not	1	2	3	4	5
The woman reports or presents with poor body image during pregnancy	1	2	3	4	5

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The woman reports or presents with low self-esteem during pregnancy	1	2	3	4	5
The woman reports (or appears to) rumination about gestational weight gain	1	2	3	4	5
The woman reports (or appears to) rumination about health of baby	1	2	3	4	5
The woman reports (or appears to) fixation on post-partum weight loss	1	2	3	4	5
The woman reports (or appears to have) self-critical thoughts and fear of criticism	1	2	3	4	5
The woman reports (or appears to) comparing personal eating habits to others	1	2	3	4	5
The woman reports (or appears to have) a need for the pregnancy to be “perfect”	1	2	3	4	5
The woman reports a desire for her baby to be “small” or “petite”	1	2	3	4	5
The woman reports (or appears to have) suicidal thoughts/ideation	1	2	3	4	5
The woman reports or presents with distress regarding her changing weight/shape/size and a fear of fatness	1	2	3	4	5
The woman reports feeling distressed or guilty after eating “unhealthy” or “bad” foods	1	2	3	4	5
The woman reports or presents with mood disturbance	1	2	3	4	5
The woman reports or presents with anxiety about certain foods/food groups	1	2	3	4	5
The woman reports feeling “out of control” of her body	1	2	3	4	5
The woman reports feeling a “loss of control” over eating	1	2	3	4	5
The woman reports or presents with guilt after eating (any food)	1	2	3	4	5
The woman reports or presents with feelings of shame and disgust about her pregnancy body/shape/size	1	2	3	4	5
The woman reports sensitivity to comments regarding pregnancy weight, shape, or appearance	1	2	3	4	5
The woman reports or presents with emotional detachment from pregnancy	1	2	3	4	5
The woman reports or appears to be socially isolated	1	2	3	4	5
The woman reports interpersonal mistrust, particularly related to her partner’s perception of her pregnancy weight/shape/size and cooking	1	2	3	4	5

If you have any comments or recommendations for additional signs or symptoms, please list these below so they can be added to Round II for consideration.

_____

_____

_____

_____

SECTION 2 – POSSIBLE DISTINGUISHING FACTORS

Below is a list of possible factors that may be used to distinguish disordered eating symptomatology in pregnancy from normative pregnancy-related changes. Please indicate how important you believe each factor is in distinguishing disordered eating symptomatology from pregnancy-appropriate symptomatology (foci items) on a 5-point Likert scale (1 = *not important* to 5 = *very important*).

	Not important	Low importance	Neutral	Important	Very important
Severity of behaviours	1	2	3	4	5
Severity of cognitions	1	2	3	4	5
Frequency of behaviours	1	2	3	4	5
Frequency of cognitions	1	2	3	4	5
Dietary behaviours in excess to recommended guidelines	1	2	3	4	5
Dietary behaviours in deficit to recommended guidelines	1	2	3	4	5
Exercise behaviours in excess to recommended guidelines	1	2	3	4	5
Exercise behaviours in deficit to recommended guidelines	1	2	3	4	5
Appropriateness of gestational weight gain	1	2	3	4	5
Health risk or distress to fetus	1	2	3	4	5
Health risk or distress to mother	1	2	3	4	5
Distress of (or worry by) family	1	2	3	4	5
History of pregnancy complications (e.g., miscarriage, premature labour)	1	2	3	4	5
Level of physical impairment or impact	1	2	3	4	5
Level of psychological impairment or impact (i.e., affective state of mother)	1	2	3	4	5
Level of social impairment or impact	1	2	3	4	5
Level of relational impairment or impact	1	2	3	4	5
Degree of flexibility with dietary rules	1	2	3	4	5
Level of insight and/or denial	1	2	3	4	5
Discrepancy between self-reported functioning and medical observations	1	2	3	4	5
Discrepancy between the woman's report and partner/family reports	1	2	3	4	5
Available coping strategies (e.g., emotion regulation skills)	1	2	3	4	5
Available social support	1	2	3	4	5
History of any psychiatric condition	1	2	3	4	5
History of an eating disorder	1	2	3	4	5

	Not important	Low importance	Neutral	Important	Very important
History of subclinical eating disorder symptoms	1	2	3	4	5
Family history of an eating disorder	1	2	3	4	5
Younger age (< 30 years)	1	2	3	4	5
Older age (> 30 years)	1	2	3	4	5
Ethnicity	1	2	3	4	5
Primigravity (first pregnancy)	1	2	3	4	5
Multigravidity (subsequent pregnancies)	1	2	3	4	5

If you have any comments or recommendations for additional distinguishing factors, please list these below so they can be added to Round II for consideration.

_____

_____

_____

_____

SECTION 3 – ASSESSMENT OF DISORDERED EATING IN PREGNANCY

This section is divided into two subsections. In the first section you will be asked to indicate your beliefs about screening for subclinical eating disorder symptoms in antenatal care. In the second section, you will be asked to review and rate the suitability of 12 potential assessment methods for identifying subclinical eating disorder symptoms in antenatal care.

Please indicate your beliefs regarding the nature of screening for subclinical eating disorder symptoms in pregnancy:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Screening for disordered eating should be a routine component of antenatal care (i.e., occur for every woman)	1	2	3	4	5
Screening for disordered eating in antenatal care should only occur when indicated by presenting signs/symptoms and/or historical factors	1	2	3	4	5
Screening for disordered eating and EDs should not occur in antenatal care	1	2	3	4	5

Please indicate how suitable you consider the following methods in assessing/identifying subclinical disordered eating symptoms in pregnancy:

	Not suitable at all	Somewhat unsuitable	Neutral	Somewhat suitable	Very suitable
Visual observation	1	2	3	4	5
Physical examination of the woman/mother	1	2	3	4	5
Fetal examination (e.g., ultrasound)	1	2	3	4	5
Pathology examination	1	2	3	4	5
Review of medical records	1	2	3	4	5
Direct questioning (e.g., *Do you have an eating disorder?)*	1	2	3	4	5
Collection of collateral information from support network (e.g., partner, family)	1	2	3	4	5
Opportunistic questioning by clinician (unstructured)	1	2	3	4	5
Brief clinician administered screening (e.g., SCOFF in an oral format)	1	2	3	4	5
Patient completed screening measures (e.g., SCOFF in a paper-pencil format)	1	2	3	4	5
Self-report questionnaires (e.g., EDE-Q, EAT, EDI)	1	2	3	4	5
Structured clinical interviews (e.g., EDE)	1	2	3	4	5

If you have any comments or recommendations for additional assessment methods, please list these below so they can be added to Round II for consideration.

_____

_____

_____

_____

Thank you.

That's the end of Round I. If you have any questions, please contact Amy at abannaty@bond.edu.au. The Round II questionnaire (including Round I feedback) will be sent in a couple of weeks.

Appendix C

Psychometric Instruments Identified During Full-Text Review in Study 3

As shown in Table 42, sixteen different instruments were identified during the full-text review process including three clinician-administered structured clinical interviews, 11 self-report questionnaires (or derived subscales/questions), one self-report screening instrument, and a series of non-standardised questions mostly from the MoBa cohort study.

Table 42

Overview of Psychometric Instruments Revealed During Full-Text Screening

Psychometric Instrument	Format	Studies
Eating Disorders Examination (EDE; Cooper & Fairburn, 1987)	Interview	• Fairburn et al. (1994) – modified EDE • Foster et al. (1996) – shape concern questions only • Conti et al. (1998) – modified EDE • Abraham et al. (2001) – modified EDE • Emery et al. (2017) – pregnancy EDE version* • Kolko et al. (2017) – pregnancy EDE version*
Structured Clinical Interview for DSM (SCID; First, Spitzer, Gibbon, & Williams, 1999)	Interview	• Quispel et al. (2015) – DSM-IV-TR version • Santos et al. (2016) – DSM-5 version ED module only* • Santos et al. (2017) – DSM-5 version ED module only*
Eating Disorder Longitudinal Interval Follow-Up Evaluation (Keller et al., 1987)	Interview	• Blais et al. (2000)
Eating Disorders Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994, 2008)	Self report	• Senior et al. (2005) – selected items from weight & shape subscales* • Micali et al. (2007) – selected items from weight & shape subscales* • Micali et al. (2011) – selected items from weight & shape subscales* • Soares et al. (2009)** • Nunes et al. (2012)** • Allison et al. (2012) • Nunes et al. (2014)** • Squires et al. (2015) – modified • Easter et al. (2015) • Tremblay (2015) • Gonçalves et al. (2015) • Pettersson et al. (2016)

Note. * or ** in each row = same sample

Table 42 (continued)

Overview of Psychometric Instruments Revealed During Full-Text Screening

Psychometric Instrument	Format	Studies
Eating Disorder Inventory (EDI; Garner, Olmstead, & Polivy, 1983)	Self report	• Davies & Wardle (1994) - drive for thinness, body dissatisfaction, and bulimia subscales only • Gough (1998) – drive for thinness and body dissatisfaction subscales only • Behar et al. (2008)
Eating Disorder Inventory-2 (EDI-2; Garner, 1991)	Self report	• Lai et al. (2005)* • Lai et al. (2006)* • Rocco et al. (2005)
Eating Disorder Inventory-3 (EDI-3; Garner, 2004)	Self report	• Broussard (2012)
Eating Disorder Diagnostic Scale (EDDS; Stice et al., 2000)	Self report	• Easter et al. (2013)* • Easter et al. (2015)*
Eating Attitudes Test-40 (EAT-40; Garner & Garfinkel, 1979)	Self-report	• Behar et al. (2008) • Annagür et al. (2014)
Eating Attitudes Test-26 (EAT-26; Garner et al. 1982)	Self report	• Turton et al. (1999) • Santos et al. (2016) • Sumner et al. (1993)
Three Factor Eating Questionnaire (TFEQ; Stunkard & Messick, 1985)	Self report	• Gough (1998) • Allison et al. (2012) – item 51 only • Slane & Levine (2015) – restraint and disinhibition subscales only
Disordered Eating Behaviour Scale (DEBS; Muazzam & Khalid, 2011)	Self report	• Sohail & Muazzam (2012)
Quality of Life Related to Eating Disorders (QOLED; Abraham et al., 2006)	Self report	• Coker et al. (2013) • Coker & Abraham (2015)
Dutch Eating Behaviour Questionnaire (DEBQ; van Strien et al., 1986)	Self report	• Davies & Wardle (1994) – restraint subscale only
Sick Control One Fat Food Questionnaire (SCOFF; Morgan, Reid, & Lacey, 1999)	Self report	• Hubin-Gayte et al. (2012) • Moalla et al. (2015)
Non-standardised study specific questions	Self report	• Bulik et al. (2007), Bulik et al. (2008), Knoph-Berg et al. (2008), Bulik et al. (2010), Dellava et al. (2011), and Zerwas et al. (2014) – MoBa specific ED questions* • Oliboni & Alvarenga (2015)

Note. * or ** in each row = same sample

Appendix D

Quality Checklist for Studies Included in Study 3 (Systematic Review)

Table 43

Assessment of Data Quality Using a Modified QUADAS Checklist and Criteria from Mirza and Jenkins (2004)

Study	Explicit aims	Sample size justification or adequate	Sample described sufficiently	Justification sample representative of population	Inclusion & exclusion criteria stated	Use of appropriate reference standard	Reliability of measure reported	Validity of measure reported	Dropouts & withdrawals explained	Data described adequately	Discussion of generalis-ability
Lai et al. (2005)*	Yes	Yes	Yes	Yes	Yes	N/A	Yes	No	Yes	Yes	Yes
Lai et al. (2006)*	Yes	No	Yes	Yes	Yes	N/A	Yes	No	Yes	Yes	Yes
Sohail & Muazzam (2012)	Yes	Yes	Yes	No	Yes	N/A	Yes	No	No	Yes	Yes
Mohamadirizi et al. (2015)	Yes	Yes	Yes	No	Yes	N/A	Yes	No	No	Yes	No
Tremblay (2015)	Yes	No	Yes	Yes	No	N/A	Yes	No	Yes	Yes	Yes
Gonçalves et al. (2015)	Yes	No	Yes	No	No	N/A	Yes	No	No	Yes	Yes
Pettersson et al. (2016)	Yes	Yes	Yes	Yes	No	N/A	Yes	No	No	Yes	Yes
Emery et al. (2017)**	Yes	Yes	Yes	No	Yes	N/A	Yes	No	No	Yes	Yes
Kolko et al. (2017)**	Yes	Yes	Yes	No	Yes	N/A	Yes	No	No	Yes	Yes

Appendix E

Ethics Approval for Study 4

BOND UNIVERSITY
BRINGING AMBITION TO LIFE

HUMAN RESEARCH ETHICS COMMITTEE
Bond University
Gold Coast, Queensland 4229
Australia
Ph: +61 7 5595 4194
Fax: +61 7 5595 1528
(from overseas)
Email: ethics@bond.edu.au

ABN 88 010 694 121
CRICOS CODE 00017B

21 April 2017

Peta Stapleton
Faculty of Society and Design
Bond University

Dear Peta

Application ID:	**15964**
Project Title:	**Disordered Eating in Pregnancy: The Development and Validation of a Pregnancy-Specific Screening Instrument**
Researchers:	**Peta Stapleton, Amy Bannatyne, Bruce Watt, Kristen MacKenzie-Shalders**

I am pleased to confirm that your project was reviewed by Bond University Human Research Ethics Committee and you have been granted approval to proceed.

The Committee requires, as a condition of approval, that all investigations be carried out in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007). Approval is subject to conduct of the research in accordance with the requirements set out in the National Statement.

Approval is given subject to the protocol of the study being undertaken as described in your application, and approved amendments. As you may be aware the Ethics Committee is required to annually report on the progress of research it has approved. We would greatly appreciate if you could respond promptly and fully to the request for information on this project which will be distributed in March/April each year.

Under the terms of the National statement BUHREC has a role to monitor approved research projects and if necessary may withdraw approval. Conduct of unapproved research or deviation from the approved protocol may constitute academic misconduct and will be investigated in accordance with Part B of the *Australian Code for the Responsible Conduct of Research* (2007). Please refer to the Research Ethics website for more detail on Research Integrity and Bond University processes for dealing with instances of research misconduct.

You are reminded that the Principal Investigator must immediately report anything that might warrant review of ethical approval of the project. Should you have any queries or experience any problems, please contact us promptly.

We wish you well with your research project.

Yours sincerely

Dr Mark Bahr
Chair Bond University Human Research Ethics Committee

Appendix F

The Disordered Eating Attitudes in Pregnancy Scale (DEAPS)

Over the past month…	**Yes**	**No**
1. I have felt distressed about the changes to my body and/or eating habits during pregnancy	☐	☐
2. I have attempted to stop the changes occurring to my body during pregnancy	☐	☐
3. I have felt anxious *about* eating in general, or about eating certain foods	☐	☐
4. I have felt distressed *after* eating because of its effect on my weight and shape	☐	☐
5. I have noticed that what I allow myself to eat and how much I can eat is connected to rules and conditions	☐	☐
6. I felt like there were times when I lost control over my eating and/or body	☐	☐
7. I worried that I have, or will, become 'fat' during pregnancy	☐	☐
8. I have felt distressed and uncomfortably full (i.e., like I am going to burst) after eating a large amount of food	☐	☐
9. I have spent considerable time researching the most effective ways to minimise how much weight I gain while pregnant	☐	☐
10. I have spent considerable time researching how I can rapidly lose weight after I have given birth	☐	☐
11. I have felt disgusted or ashamed with my pregnancy body	☐	☐
12. My evaluation of my body shape, weight, or size during pregnancy has significantly influenced how worthy I believe I am as a mother or person	☐	☐
13. I have wanted my pregnancy body to be small, like I am "just bump" (i.e., only my stomach appears to have grown, with no weight or shape changes to other areas of my body)	☐	☐
14. I have found myself frequently (at least once a week) comparing my weight, shape, size, or eating habits to other women	☐	☐
TOTAL	_____	

Scoring Notes. Assign a score of 0 to 'no' responses and a score of 1 to 'yes' responses. A total score equal to or greater than 8 represents a positive screen for possible disordered eating attitudes. Further monitoring and assessment is required. This may include referral to specialist services if clinically indicated.